Evolution Processes Information –

Linking the Theory of Evolution and Information Theory

Martin Hilbert

Department of Communication, University of California, Davis, CA, 95616, USA

Correspondence to: hilbert@ucdavis.edu

The article reformulates evolutionary fitness as information processes. This changes our understanding of evolutionary change, as it allows its formal expression as a communication process between the evolving system and its environment. Similar to how a communication process reveals a communicated message among possible messages, evolution reveals the surviving fittest member of a population. The amount of uncertainty reduced by evolution is quantified by 'negentropy', which turns into relative entropy and mutual information in the general cases. Bits and bytes become a quantifiable ingredient of evolutionary growth. More information between the population and its environment implies greater fitness. Information also turns out to be the link between the distinct population types in space and the unfolding environmental patterns in time.

ABSTRACT

Recent studies have shown that specific aspects of evolutionary dynamics can be expressed in information theoretic terms. This article connects and generalizes these partial results by presenting several complete multilevel decompositions of evolutionary fitness in terms of information processes. We formally show how evolution produces negative entropy, which turns into Kullback-Leibler relative entropy in the general case, and into mutual information in the case of independence between the evolving population and its environment. We encounter information theoretic ways to express and link longstanding concepts of evolutionary theory, such as Malthusian fitness, Fisher's fundamental theorem of natural selection, the Price equation, the replicator equation, evolutionary sustainability, and multilevel emergence. This leads to new conceptualizations of well-known phenomena, such as the expression of changes in fitness as the population's descriptive complexities from the perspectives of the past and future. Fitness itself attains an intuitive, but formal interpretation as the amount of informational 'fit' between the evolving population and its environment, expressed as mutual information. We also find a space-time version of Ashby's law of requisite variety between the evolving population and the corresponding environmental pattern, and a generalization of Kelly's bet-hedging result, both time-honored results from 1956. The insights gained from the reformulations of fitness as evolving information underline the familiar argument that scientific advancements often simply draw from ontological re-conceptualizations of well-known problems. Showing the generality of the results, the obtained theory is applied to calculate the amount of informal bits processed by several dynamics of socio-economic and cultural evolution.

While it is often suggested that evolution can be interpreted in terms of an information process that transforms complexity between the environment and the evolving system *(1)*, a more formal treatment of this idea is still incomplete. Recently important advancements have been made in the establishment of formal links between the theory of evolution and information theory *(2-8)*. Among them, it has been shown that the part of the evolutionary process that refers to natural selection (defined by Fisher's fundamental theorem as the variance in growth rates *(9-12)* turns out to be equivalent to the information theoretic measure of Fisher information *(2,3,12-14)*. While these results provide support that evolutionary aspects can be understood in terms of evolving information, they exclusively focus on natural selection. As the evolutionary theorist Ronald Fisher reminded us with the first sentence of his influential 1930 book: "Natural Selection is not Evolution" *(9; p.vii)*, but merely part of the evolutionary dynamic *(10,15)*. This article presents several decompositions of complete evolutionary change in terms of information theory (in the sense of *(16-18)*), and links them to longstanding concepts that have turned out to be useful when reasoning about evolutionary change. Being descriptive of any kind of evolutionary process, it is not subject to limiting modelling assumptions. Showing the generality of the results, the decompositions are applied to several empirical cases of human socio-economic and cultural evolution, including the evolution of exports in an economy, the evolution of fundraising dynamics of entrepreneurs, and the evolution of online social media.

Overview

The first section starts the exploration by expressing evolutionary fitness in terms of entropies. The concept of entropy has filled the imagination of scientists since the 1800s, and is one of the most fundamental metrics in physics *(19-21)*, statistics *(22,23)* and communication *(16,18)*. It basically measures the level of surprise or uncertainty in a distribution. Evolution can be

understood as a probability game that resolves uncertainty with regard to the 'survival of the fittest (type)'. The resolution of uncertainty can be formally presented as negative entropy, or 'negentropy' (equations (1.2) ad (2.1)).

The second section generalizes this logic to evolutionary processes that are out of equilibrium and still ongoing (no determination of the 'fittest' (yet)). Relative entropy turns out to be the adequate information theoretic metric in this case. Being a relative metric, it allows to quantify the resolved uncertainty relative to the present- or relative to the future generation. From the perspective of the future, uncertainty has already been resolved, resulting in the subtraction of relative uncertainty, while from the perspective of the present, the descriptive complexity of the involved uncertainty (such as expressed in informational bits) is a positive ingredient of fitness, and is still to be exploited (equations (3.1) ad (3.2)).

The third section shows that information theoretic formulations of evolutionary change naturally lend themselves to recursive multilevel decompositions that consider that evolution processes information on different levels (equation (4.1)). It turns out that the resulting conditional descriptive complexities over different levels add up to a total that is equal to the descriptive complexity of the evolving individuals (equation (4.2)). One of the practical applications of the presented concepts calculates the amount of information processed by the Canadian export economy, classified over six taxonomic levels of economic activity. Just like a lion is a cat, which is a type of carnivore, which evolves among all mammals, which are types of vertebrates of the animal population, this example from economic evolution considers that the export product 'color TV' (code 7611 of UN SITC, rev.2) evolves within other types of 'TV receivers', which is part of the higher level type of 'telecommunication equipment', which evolves among 'machinery equipment', which is a type of 'machinery' that belongs to the population of all

'manufacturing' products. Fig. 1 shows the varying amounts of descriptive complexities in bits produced by evolution over four decades on each of these levels.

The fourth section explores a variety of applications of the proposed logic to partial and more general aspects of evolutionary change (such as natural selection and changes in fitness). For example, it turns out that the change in fitness in a stable setting can be expressed as the descriptive complexity of the present population from the perspective of the past, plus the descriptive complexity of the present from the perspective of the future (equation (5)). This says that evolutionary acceleration is equal to present complexity looked at from the past and future. The rushed reader can skip the explorations of this fourth section in a first reading without prejudice of understanding the final section.

The final fifths section expands the logic from the first three sections to include the environment. What took the form of entropy in the first section, and relative entropy in the general case of the second section, now takes the form of mutual information between the evolving system and its environment (equations (8.4) and (8.5)). 'Fit-ness' attains an intuitive, but formal interpretation as the amount of 'fit' between the evolving population and its environment. Fitness is related to the amount of informational 'fit' on a certain level. This amount of mutual information represents the emergent quantity that makes the total population fitness on a higher level more than the sum of the fitness of its parts on a lower level. Natural corollaries from this way of looking at evolution lead to generalizations of two important results from 1956, namely Kelly's bet-hedging strategy (equation (8.6)), and Ashby's law of requisite variety between an evolving population and its environment (equations (9.1) and (9.2)).

The concluding discussion reviews the main argument, which is that the expression of evolutionary dynamics in terms of informational descriptive complexities leads not only to fresh ways of looking at well-known dynamics of evolutionary change, but also to meaningful new results.

Fitness as negative entropy

We start our explorations with the traditional interpretation of fitness w as the factor of reproduction or growth, and define the total number of population units n at time t with N^t, resulting in the population fitness: $w = \left[\frac{N^{t+1}}{N^t}\right]$. The units can represent the number of current individuals N^t and future offspring N^{t+1}; pennies of gamble bets N^t and payoff N^{t+1}, etc. We start by assuming that each individual n is its own unique type (e.g. no two finches have exactly the same peak length, so for now we assume all present and future finches to be individual types) and that each unique type has the same evolutionary weight, $p(n) = \frac{1}{N}$ (e.g. one finch represents the same population proportion as another finch), which results in a uniform probability mass function among them $\sum_{n^t} p(n^t) = N^t * \frac{1}{N^t} = 1$. Without loss of generality, we can represent growth factors on a logarithmic scale, which is often referred to as Malthusian fitness *(9,24)*. This essentially normalizes fitness at 0 for an unchanging population, $\log(w = 1) = 0$. For example, a logarithm of base 2 represents fitness w in terms of the number of population doublings at each time step. This rudimentary set up is sufficient to express fitness in terms of Shannon entropies *(16,18)*:

$$\log(w) = \log\left(\frac{N^{t+1}}{N^t}\right) = \log(N^{t+1}) - \log(N^t) = E^{t+1}[\log(N^{t+1})] - E^t[\log(N^t)]$$

$$= -\sum_{n^{t+1} \in N^{t+1}} p(n^{t+1}) \log\left(\frac{1}{N^{t+1}}\right) + \sum_{n^t \in N^t} p(n^t) \log\left(\frac{1}{N^t}\right)$$

$$= H(N^{t+1}) - H(N^t) \qquad (1.1)$$

$E^{t+1}[...]$ and $E^t[...]$ stand for the expected values at time $t+1$ and t, respectively, and the second reformulation makes use of the fact that the expected value of a constant is the constant. Equation (1.1) shows that if we assume that each individual of the evolving population has its own, uniquely distinguishable type, fitness measures the uncertainty inherent in the future population, minus the uncertainty of the present population (both consisting with different "alphabets", or "number of types" in this case, which can be due to minuscular mutations, etc.). In the case that mutations that result in many more unique types at step $(t+1)$, uncertainty increases, and $H(N^{t+1}) \gg H(N^t)$ (see equation (1.1)). In the case that evolution leads to the survival of one single member at the final equilibrium stage at time T ('survival of the fittest'), there is no uncertainty anymore about who of all population members is the chosen one and $H(N^{t+1}) = H(N^T) = 0$. Therefore, the expression of an evolutionary fitness over the period $0 \to T$ from the initial population at $t = 0$ until the exclusive survival of the fittest at time T is:

$$\log(w^{0 \to T}) = -H(N^t) \qquad (1.2)$$

The negativity of entropy (sometimes also referred to as "negentropy" *(25, 26)*) shows that evolution --just like any process of communication *(16,18)*-- reduces uncertainty. While a communication processes reveals a desired symbol of an uncertain alphabet, evolution can be understood as a process that reveals the fittest member of a population. The amount of uncertainty reduced by evolution is measured by the amount of negative entropy $H(N^t)$.

Normally, evolutionary theory does not consider each individual as one unique type with equal weight, but aggregates individuals into groups that are distinguishable and justifiable by some common criteria called 'type' (e.g. a certain group of all 'long-peaked finches' and another group of all 'short-peaked finches'), whereas the weight of types change as the population 'evolves'. Groups of a certain type can also be aggregated into even lower level groups of a common subtype, leading to a multilevel logic that is essential to evolutionary thought. As shown in this article, multilevel representations are a natural part of information theoretic reformulations of evolutionary dynamics, which is the reason why we need a scalable representation of levels.

We define g_l as a group that joins a certain number of units n under a common type. The coarse-graining is taken on level l and G_l acts as a discrete random variable with a probability mass function of types $\sum_{g_l} p(g_l) = 1$ (for a more detailed introduction to the notation used in this article, please see Supporting Material (S.1)). On the highest level we count with only one group (the entire population), $p(g_{l=0} = population) = 1$, with population fitness $w(g_0) = \left[\frac{N^{t+1}}{N^t}\right]$ (often referred to as $\bar{w}$ in the nomenclature of evolutionary biology). Each group counts with a collective fitness defined as the weighted average fitness of its members on the next lower level:
$w(g_l) = \sum_{g_{l+1}} p(g_{l+1}) * w(g_{l+1}) = E[w(g_{l+1})]$, its expected value.

In the case that evolution leads to the survival of one single type at time T ('survival of the fittest type'), we get (see Supplementary Material (S.2.1.4)):

$$\log\left(w(g_{l=0}{}^{t \to T})\right) = \log w(gmax_{l=1}) - H(N^t) \qquad (2.1)$$

whereas $w(gmax)$ stands for the highest fitness among all types during the period $t \to T$ (the 'fittest' individual member of the population). In words, equation (2.1) says that population

fitness over the process of evolution until 'the survival or the fittest type' is equal to the fitness of the fittest type minus the uncertainty that was originally inherent in the initial population.

Fitness as relative descriptive complexity

We now generalize the previous logic to the case that evolution has not (yet) settled down at an equilibrium of the fittest type and is still ongoing. The natural information theoretic metric for out of equilibrium dynamics turns out to be the Kullback-Leibler divergence *(17)* between the future and the present populations (see Supplementary Material (S.1.1) for an illustrative example and (S.2.1) for the complete derivation).

$$\log(w(g_l{}^t)) = \sum_{g_{l+1}{}^{t+1}} p(g_{l+1}{}^{t+1}) * \log\left(w(g_{l+1}{}^t) * \frac{p(g_{l+1}{}^t)}{p(g_{l+1}{}^{t+1})}\right)$$

$$= \sum_{g_{l+1}{}^{t+1}} p(g_{l+1}{}^{t+1}) \log(w(g_{l+1}{}^t)) - \sum_{g_{l+1}{}^{t+1}} p(g_{l+1}{}^{t+1}) \log\left(\frac{p(g_{l+1}{}^{t+1})}{p(g_{l+1}{}^t)}\right)$$

$$= E^{t+1}[\log w(g_{l+1}{}^t)] - D_{KL}(P^{t+1}||P^t) \qquad (3.1)$$

$$\log(w(g_l)) = E^t [\log w(g_{l+1})] + D_{KL}(P^t||P^{t+1}) \qquad (3.2)$$

The involved reformulation employs a reversed form what is known as the 'replicator equation' *(27)*: $p^{t+1} = p^t * \frac{w}{E[w]}$ (see Supplementary Material (S.2.1)). D_{KL} is also known as relative entropy *(18)* and is an asymmetric measure of divergence that quantifies the directed increase in descriptive complexity if the distribution P^t is encoded with the optimized code of the distribution P^{t+1} or vice versa (for most cases with $D_{KL}(P^{t+1}||P^t) \neq D_{KL}(P^t||P^{t+1})$). For example, the optimal code for the distribution P^{t+1} has average description length of $H(P^{t+1})$, but encoding P^{t+1} with the optimal code for P^t increases the descriptive complexity by $D_{KL}(P^{t+1}||P^t)$. The sum of $H(P^{t+1}) + D(P^{t+1}||P^t)$ is the so-called cross-entropy $H(P^{t+1}, P^t)$,

which (assuming logarithm of *base 2*) is the descriptive complexity in the number of *bits* required to encode the updated population with a code based on the original population. Thus, D_{KL}, such as H, is a measure of uncertainty and is always positive (result of the concavity of the logarithmic function *(18; p.28))*.

In words equation (3.1) says that current population fitness is equal to the average fitness of the future's individuals, minus the uncertainty of the future population at time $t + 1$ when looked at from the perspective of the present at time t. In this case, the relative entropy $-D_{KL}$ takes the same form as the absolute entropy H from equations (1.2) - (2.1). Equation (3.2) states that current population fitness is equal to the average fitness of the current individuals at time t, plus the uncertainty still to be exploited by the ongoing evolutionary change when looking at the present from the perspective of the uncertain future at $t + 1$. Since $D_{KL} \geq 0$, the formulation of equation (3.2) also gives us a formalization of multilevel emergence, whereas the total is more than the (weighted) sum of its parts: $\log(w(g_l)) \geq \sum p(g_{l+1}{}^t) * [\log w(g_{l+1})]$. In words, the total population fitness on the higher level l is more than the sum of the fitness of its members on the lower level $l + 1$. How much more? Exactly $D_{KL}(P^t \| P^{t+1})$ more. Thus, multilevel emergence can be quantified in terms of the involved relative uncertainty. In mathematical terms, this result quantifies the inequality involved in the logarithmic case of Jensen's inequality *(28)*:

$$\log(E[w(g)]) \geq E[\log(w(g))].$$

Multilevel evolution

We can expand this two level logic to a multilevel logic. Just like biological life is traditionally classified into eight major taxonomic ranks (domain, kingdom, phylum, class, order, family, genus, species) *(29)*, an economy consists of economic sectors, industries, divisions, product

groups, and products *(30-32)*. Any such taxonomy assumes that a certain type at level $l = 1$ consists of an underlying distribution of subtypes, each with a certain distribution of subtype fitness on the next lower level $l = 2$. As such, we can structure a population into groups by fine-graining types into subtypes, resulting in a multilevel logic over different levels from the highest level population level $l = 0$, until the lowest most fine-grained level permitted by the database, denoted by $l = L$, and therefore: $l = \{0,1,2 \ldots L\}$. The setup of equation (2.1), (3.1) or (3.2) reveals a recursive multilevel logic whereas we can decompose $\log(w(g_{l+1}))$ on the right hand side, with the same logic we decomposed $\log(w(g_l))$ on the left hand side (much in the sense of Price's well-known multilevel decomposition of evolutionary change *(33,34)*). The following notation represents the fact that the lower level group is nested within the higher level type by a reversed slash $' \backslash '$, whereas the proportions of the types are normalized on the higher level to always sum up to 1, just like when working with conditional probabilities *(35)*: $p(g_{l=0}) = 1; \sum p(g_{l=1}) = \sum p(g_{l=1} \backslash g_{l=0}) = 1; \sum p(g_{l=2} \backslash g_{l=1,0}) = 1$. For example, decomposing the case of equation (3.2) into three levels (population: $l = 0$; types: $l = 1$; individuals n: $l = 2 = L$), we get:

$$\log w(g_{l=0}) = \sum_{g_{l=1}} p(g_{l=1})\{\log w(g_{l=1})\} + D_{KL}(P(g_{l=1})^t \| P(g_{l=1})^{t+1})$$

$$= \sum_{g_{l=1}} p(g_{l=1}) \left\{ \sum_{g_{l=2}} p(g_{l=2}\backslash g_{l=1}) \log(w(g_{l=2})) \right.$$

$$+ \sum_{g_{l=2}} p(g_{l=2}\backslash g_{l=1})^t \log \left(\frac{p(g_{l=2}\backslash g_{l=1}{}^t)}{p(g_{l=2}\backslash g_{l=1}{}^{t+1})} w(g_{l=1}) \right) \Bigg\}$$

$$+ D_{KL}(P(g_{l=1})^t \| P(g_{l=1})^{t+1})$$

$$= \sum_{g_{l=2}} p(g_{l=2})^t \log(w(g_{l=2})) + D_{KL}(P(g_{l=2}|g_{l=1})^t \| P(g_{l=2}|g_{l=1})^{t+1})$$

$$+ D_{KL}(P(g_{l=1})^t \| P(g_{l=1})^{t+1}) \qquad (4.1)$$

$$= \sum_{g_{L=2}} p(g_{L=2})^t \log(w(g_{L=2})) + D_{KL}(P(g_{L=2})^t \| P(g_{L=2})^{t+1}) \qquad (4.2)$$

The step from equation (4.1) to (4.2) makes use of the chain rule for relative entropy (*(18)*; p. 24; see also *(35)*). We can always follow this logic through from the highest level to the lowest level $P(g_L)$ of any multilevel population structure. The sum of the conditional relative entropies over all levels (4.1) is equal to the relative entropy on the lowest, most fine-grained level (4.2). The form of equation (4.1) expresses evolutionary fitness as a sum of relative entropies from different levels (their contribution may vary at each level) and a residual term which cannot be further decomposed on basis of the available data. Equation (4.2) has the same form as equation as equation (3.2), just on the lowest level of fine-graining L permitted by the database. In practice, the equality between equations (4.1) and (4.2) allows for a useful shortcut when calculating the total amount of information processed by an evolutionary multilevel dynamic. It is straightforward to apply this chain rule logic to more than three levels (with maximally $\log_2(N)$

levels, since the smallest possible group size at each level is 2), as well as to most of the
equations presented in this article (i.e. equations (2.1), (3.1), (7.2), (8.3-8.6)).

Fig. 1 provides an empirical application this logic to the evolution of exported products (in US$)
(Fig. 1a); and the evolution of the number of views of online news video clips (Fig. 1b). It shows
over which time period which levels adds or subtracts how much descriptive complexity
(calculated in bits). Both figures show that relative entropy tends to be larger at lower levels,
which is due to the greater shifts in population constellations on less aggregate levels, but at the
same time, Fig. 1b makes clear this is not inevitably the case. Fig. 1b contrasts the descriptive
complexity measured from the present, $D_{KL}(P^{t+1}||P^t)$ (striped: multilevel form of equation
(3.1)), with the descriptive complexity measured from the future, $D_{KL}(P^t||P^{t+1})$ (filled:
multilevel form of equation (3.2)). It shows that D_{KL} stays at the same order of magnitude among
levels, while the round 'zoomed-in inserts' confirm the asymmetry of the D_{KL} metric within the
respective order of magnitude.

Exploring the extremes of these kinds of decompositions, it turns out that population fitness can
be completely decomposed into descriptive complexities in the form of relative entropies if the
weighted geometric mean fitness of its types is 1: $\prod_{g_L} w(g_L)^{p(g_L)} = 1 =>$

$\sum_{g_L} p(g_L)\{\log w(g_L)\} = 0$ and (taking the form of equation (4.2)):

$$\log w(g_{l=0}) = D_{KL}(P(g_L)^t||P(g_L)^{t+1}) \qquad (4.3)$$

In the case that $P(g_L)$ is uniform (e.g. all equally weighted individuals, with $p(g_L) = p(n) = \frac{1}{N}$),
this implies that the combinatory potentials of present and future generations are equal: $\prod_{n^t} n^t =$

$\prod_{n^{t+1}} n^{t+1}$ (given that $w(g_L) = \frac{n^{t+1}}{n^t}$, with n being the number of units of a certain group g_L,

solve for $\sum \frac{1}{N^t} \left\{ \log \left(\frac{n^{t+1}}{n^t} \right) \right\} = 0$). On the other extreme, it shows that evolution does not process information if there is no change in the population constellation: with $w(g_l) = $ constant for all types, $P^t = P^{t+1}$, and $D_{KL}(P^t || P^t) = 0$. This underlines the argument that the parameter that matters in evolutionary dynamics is not the passing of time per se, but perceivable change of the population constellation (compare the assumption of (14), with the more sophisticated argument made in (2,3)).

Further information-theoretic reformulations

This section shows that the expression of fitness in information theoretic terms can be expanded to look at a variety of specific aspects of evolution, such as natural selection, and at more general aspects, such as the change in fitness or long-term fitness.

We start by asking about the change Δ in fitness (such as done Price's famous equation (33,34)). Equation (5) shows that for stable individual fitness on the lower level, $w(g_{l=1}{}^{t-1}) = w(g_{l=1}{}^t)$, the change in population fitness on the higher level is equal to the descriptive complexity of the present system when looked at it from the perspective of the past, plus the descriptive complexity if the present system when looked at it from per perspective of the future (see also (S.2.3.2)):

$$\log w(g_{l=0}{}^t) - \log w(g_{l=0}{}^{t-1}) = \log \frac{w(g_{l=0}{}^t)}{w(g_{l=0}{}^{t-1})} = D_{KL}(P^t || P^{t-1}) + D_{KL}(P^t || P^{t+1}) \qquad (5)$$

This throws a new light on a well-known concept. In a stable setting we can express the concept of changing fitness entirely in terms of the descriptive complexities of the present moment (looked at from the perspectives of the past and future). The positivity of D_{KL} ((18); p.28) confirms an inevitable positive acceleration of population growth in a stable setting.

14

This is in line with Fisher's fundamental theorem of natural selection, which measures natural selection in terms of variance in fitness *(9-12)*, a term that is also always positive. Same as variance, relative entropy measures diversity in fitness: the more diversity in $w(g_l)$, the more distinct $P(g_l)^t$ and $P(g_l)^{t+1}$, and the larger D_{KL}. The symmetric symmetric version D_{KL} (Jeffreys divergence: $J(P^t, P^{t+1}) = D_{KL}(P^t \| P^{t+1}) + D_{KL}(P^{t+1} \| P^t)$ *(36)*) has a direct relationship to the variance in fitness, which can be shown by the well-known Price equation *(33,34)*. The following replaces the freely selectable parameter z_i of Price's setup with logarithmic fitness $\log(w(g_l))$ and asks for the change in log fitness (for the derivation see S2).

$$w(g_0)\Delta z = Cov(w(g_1), z_i) + E[w(g_1) * \Delta z_i] \qquad (6.1; \ Price \ equation)$$

Insert $z_i = \log(w(g_{l=1}))$ and $\Delta z = (\log w(g_{l=1}{}^{t+1}) - \log w(g_{l=1}{}^t)) = \log w \left(\frac{g_{l=1}{}^{t+1}}{g_{l=1}{}^t} \right)$

$$\frac{Cov(w(g_1), \log w(g_1{}^t))}{w(g_0)} = D_{KL}(P(g_1)^{t+1} \| P(g_1)^t) + D_{KL}(P(g_1)^t \| P(g_1)^{t+1})$$

$$= J(P(g_1)^t, P(g_1)^{t+1}) \qquad (6.2)$$

Frank *(3)* had previously proposed a proportionality between natural selection's $\frac{VAR(w(g_1))}{w(g_0)}$ and Jeffreys divergence. Equation (6.2) fine-tunes this result. While $\frac{VAR(w(g_1))}{w(g_0)} = \frac{Cov(w(g_1), w(g_1))}{w(g_0)}$, it shows that $J(P(g_1)^t, P(g_1)^{t+1}) = \frac{Cov(w(g_1), \log w(g_1{}^t))}{w(g_0)}$. The monotone increasing nature of the logarithmic function leads to the proposed proportionality, as well as to the fact that $\frac{Cov(w(g_1), \log w(g_1{}^t))}{w(g_0)}$, just like $J = D_{KL} + D_{KL}$, is always positive (*(18)*; p.28), just like the variance in fitness in Fisher's fundamental theorem.

Supplementary Material (S.2.3) explores other possible decompositions that arise from the logic presented. Several of them might merely be playful mathematical equivalences. Which of these possible reformulations turn out to be useful to obtain insights into the dynamics of evolution? One thing we know about evolution is that it takes a long time, so we could explore what we get when decomposing long term growth rates from generation t, over $t + 1$, to generation $t + 2$.

$$\log w(g_l^{t,t+1}) = \log\big(w(g_l^t) * w(g_l^{t+1})\big) \qquad (7.1)$$

$$= E^{t+1}\big[\log\big(w(g_{l+1}^t) * w(g_{l+1}^{t+1})\big)\big] - D_{KL}\big(P_{l+1}^{t+1}\|P_{l+1}^t\big) + D_{KL}\big(P_{l+1}^{t+1}\|P_{l+1}^{t+2}\big) \quad (7.2)$$

In words equation (7.2) says that the long-term population fitness over two periods is equal to the average fitness of the types, minus the descriptive complexity looked at from the original starting state, plus the descriptive complexity looked at from the final state. In this sense, equation (7.2) evaluates fitness at an intermediate point of a partitioned dynamic at which some uncertainty of the original state has already been resolved $(-D_{KL})$, while some is still pending $(+D_{KL})$.

Applying a recursive logic of multilevel decomposition in the spirit of equation (4) to equation (7.2), Fig. 2 shows the evolution of US\$ fundraised by micro-entrepreneurs structured over seven levels. It turns out that in this empirical case, $[D_{KL}(P^{t+1}\|P^{t+2}) - D_{KL}(P^{t+1}\|P^t)] \leq 0$. As shown in Supplementary Material (S.2.4), this implies that there is a sustainable positive correlation between current fitness and long-term fitness, $Cov\big(w(g^t), w(g^{t,t+1})\big) \geq 0$. This positive covariance between current and long-term fitness indicates that selection forces of the present work into the same direction of population change than the long-term forces of selection. This means that if the solved uncertainty $D_{KL}(P^{t+1}\|P^t)$ is larger than the remaining uncertainty $D_{KL}(P^{t+1}\|P^{t+2})$ within a long-term dynamic from t to $t + 2$, current selection is sustainable, because current fitness $w(g^t)$ is in line with the direction of long term fitness $w(g^{t,t+1})$.

Sustainability of fitness patterns can be meaningfully expressed in terms of descriptive complexities of the population. The other way around, if the covariance between the current and long term growth rate is negative, $Cov(w(g^t), w(g^{t,t+1})) \leq 0$, which means that current selection and long-term selection work in opposing directions, it follows that the descriptive complexity of the present $(t + 1)$ when looked at from the past (t) is smaller than the descriptive complexity when looked at from the future $(t + 2)$: $D_{KL}(P^{t+1}\|P^{t+2}) - D_{KL}(P^{t+1}\|P^t) \geq 0$ (see (S.2.4.3)). As a third case, if the descriptive complexities of the present from the perspective of the past and future are equal, $D_{KL}(P^{t+1}\|P^{t+2}) = D_{KL}(P^{t+1}\|P^t)$, it follows that $P^t = P^{t+2}$, which leads to the intuitive implication that either no evolution takes place on this level (no population change), or that selection in the second sub-period reverses previous selection, returning the population constellation at time $(t + 2)$ to the original state t (see (S.2.4.5)). Fig. 2 shows that the term is much closer to 0 in the second period of the empirical example. This makes sense, since it is to be expected that much less population change takes place during the later period in which fundraising efforts are already consolidated (successful entrepreneurs get closer or achieved their fundraising goal, unsuccessful ones stagnate).

Another interesting measure of sustainability of evolutionary forces is the covariance between present and future fitness: $Cov(w(g^t), w(g^{t+1}))$. This asks if the fittest types of the past are on average also the fittest types of the future. For example, this matters when evaluating the selection effectiveness of economic market forces *(37)*. Derivation (S.2.5) shows that if present and future selection work into the same direction of change, $Cov(w(g^t), w(g^{t+1})) \geq 0$, it follows that the long-term descriptive complexity is larger than the short term descriptive complexity: $D_{KL}(P^{t+2}\|P^t) \geq D_{KL}(P^{t+2}\|P^{t+1})$ and $D_{KL}(P^t\|P^{t+2}) \geq D_{KL}(P^t\|P^{t+1})$. Likewise,

if short term relative entropy is larger than long term relative entropy, $D_{KL}(P^{t+2}\|P^t) \leq$ $D_{KL}(P^{t+2}\|P^{t+1})$, it follows that current and future forces of selection work in opposing direction on the change of a population constellation: $Cov\big(w(g^t), w(g^{t+1})\big) \leq 0$. As shown in (S.2.6), for most cases this bound can even be tightened to the question if long-term descriptive complexity is larger or smaller than the sum of the included short-term complexities: $\{D_{KL}(P^{t+2}\|P^t)\} \geq$ $or\ \leq \{[D_{KL}(P^{t+1}\|P^t) + D_{KL}(P^{t+2}\|P^{t+1})]\}$. This shows that descriptive complexities do not fulfill the traditional triangle inequality, since the number of bits processed over a longer period can be larger or smaller than the sum of the number of bits processed during the constituent sub-periods.

Fitness as information between the system and its environment

The foregoing information theoretic decompositions have focused exclusively on quantifying an evolving population in terms of its descriptive complexities (such as measured in bits), and have not considered the environment so far. It seems intuitive that evolution should process information between the evolving population and its environment. It is straightforward to derive this connection from our previous explorations. This can be seen when assigning some kind of environmental distribution to the different short-term periods of $w(g_l{}^t)$ and $w(g_l{}^{t+1})$ in equation (7.1). What matters in order to obtain an informational value is that these different periods have distinguishable fitness for different types: information "is a difference which makes a difference" *(38)*. Therefore, one justifiable identification of meaningful environmental conditions consists in identifying periods in which the different types have different constellations of relative fitness (i.e. in one period one type grows faster than the average, while in the other period the other type grows faster). These distinguishable growth patterns then tell us something about differentiable environmental circumstances that affect the population. For

example, if in certain periods conifers thrive faster than flowering orchids, it could be inferred that the environment favors the former during these periods (e.g. the climate could be rather cool than hot), and vice versa.

Differential growth patterns allow us therefore to characterize a long-term period into a number of distinct and recurrently appearing environmental states e (in this case two distinct states). In the least informative case, all states occur with the same likelihood (highest entropy state of the environment). In the most informative case, one state explicitly follows the next (no (conditional) uncertainty), leading to a deterministic and learnable algorithm in time *(37)*, such as in a periodic orbit (e.g. day and night). In the in-between case, each state e appears with a non-uniform probability, without an exact sequence (such as a random walk of bull or bear markets; rainy, cloudy and sunny days, etc.). Even so the informational potential of the latter case is lower (since transitions are not deterministic but probabilistic), the environment still displays a patterns that contains information that can be exploited by the evolving population. The population fitness during a particular environmental state e is its expected average fitness (over all types) given (conditioned on) the environmental state e: $w(\overline{g_0}|e)$. The long term logarithmic population fitness over T periods can then be partitioned in time into the product of the growth factors in the distinct environmental states e, each appearing in its proportion $p(e)$ (for details see Supporting Material (S.2.7), i.e. (S.2.7.2)). For example, the overall growth rate of a forest over the entire year from 0 to T is the product of the forests growth during the share of sunny days $p(e = sunny)$, and during the share of days without sun $p(e = no - sun)$, during the year $0 \rightarrow T$.

$$\log(w(g_0{}^{0 \rightarrow T})) = \log\left(\prod_e w(\overline{g_0}|e)^{p(e)*T}\right) = \sum_e \log\left(w(\overline{g_0}|e)^{p(e)*T}\right) \qquad (8.1)$$

$$\log w(g_0{}^{0\to T})^{\frac{1}{T}} = \log w(\overline{\overline{g_0}}) = \sum_e \log\big(w(\overline{\overline{g_0}}|e)^{p(e)}\big)$$

In this notation the single-bar emphasizes the (arithmetic) 'space-average' over the different population members, while the double-bar emphasizes the additional (geometric) 'time-average' over the observed period. Following the formerly presented logic from above (i.e. (S.2.3)), we can expand and reformulate this average population fitness $\log w(\overline{\overline{g_0}})$ in space and time:

$$\log w(\overline{\overline{g_0}}) = \sum_e p(g_1{}^{t+1}|e) * \log\big(w(\overline{\overline{g_0}}|e)^{p(e)}\big) = \sum_e p(\overline{g_1{}^{t+1}}|e) * p(e) * \log w(\overline{\overline{g_0}}|e)$$

$$= \sum_e p(\overline{g_1{}^{t+1}}|e) * p(e) * \log\left(\frac{p(e)}{p(e)} * \frac{p(\overline{g_1{}^{t}}|e)}{p(\overline{g_1{}^{t+1}}|e)} w(\overline{g_1},e)\right) \quad (8.2)$$

$$= -\sum_e p(\overline{g_1{}^{t+1}},e) * \log\left(\frac{p(\overline{g_1{}^{t+1}},e)}{p(\overline{g_1{}^{t}},e)}\right) + \sum_e p(\overline{g_1{}^{t+1}},e) * \log\big(w(\overline{g_1},e)\big)$$

$$= E^{t+1,e}[\log w(\overline{g_1},e)] - D_{KL}\big(P(\overline{g_1{}^{t+1}},e)\|P(\overline{g_1{}^{t}},e)\big) \quad (8.3)$$

This brings the decomposition into the familiar form of equation (3.1), but this time also considering the informational pattern of the environment in a joint distribution between the population and the environment, e.g. $p(\overline{g_1{}^{t}},e)$. As an illustrative application of equation (8.3), Supplementary Material (S.2.7) shows how the evolution of USA's export economy during the period between 1979 and 2000 processed 0.0007 bits of joint relative entropy on the first level of fine-graining. This showcases how equation (8.3) can readily be applied to practical cases.

It is worthwhile noting that if the population at the population is independent from the environment (which is a reasonable assumption under certain circumstances), the joint relative entropy D_{KL} turns into Shannon's much celebrated metric of mutual information *(16, 18)*

between the environment and the future state of the population; i.e. if $p\left(\overline{g_1}^t, e\right) = p\left(\overline{g_1}^{t+1}\right) *$

$p(e)$), it follows that $D_{KL}\left(P\left(\overline{g_1}^{t+1}, e\right) \| P\left(\overline{g_1}^t, e\right)\right) = \sum_e p(g_1^{t+1}, e) \log\left(\frac{p\left(\overline{g_1}^{t+1}, e\right)}{p\left(\overline{g_1}^{t+1}\right)*p(\overline{e})}\right) =$

$I(\overline{G^{t+1}}; Env)$.

$$\log w(\overline{g_0}) = E^{t+1,e}[\log w(\overline{g_1}, e)] - I(\overline{G^{t+1}}; Env) \qquad (8.4)$$

Similarly, the mutual information between the environment and the present population emerges naturally from the equation. This is always the case at the final end state of an evolutionary process in which only the fittest has survived: $p\left(\overline{g_1}^{t=T}, e\right) = p\left(\overline{g_1}^{t=T} = fittest\right) * p(e) = 1 *$ $p(e)$.

$$\log w(\overline{g_0}) = E^{t,e}[\log w(\overline{g_1}, e)] + I(\overline{G^t}; Env) \qquad (8.5)$$

Equations (8.4) and (8.5) are the conceptual equivalent of equations (3.1) and (3.2). Equation (8.5) shows that mutual information between the current population and its environment adds to current population fitness on a certain level. Fitness obtains an intuitive explanation as the amount of mutual information 'fit' between the evolving system and its environment. The amount of mutual information also quantifies the amount of level-specific emergence involved in an evolutionary process: the total $\log w(\overline{g_0})$ is exactly $I(\overline{G^t}; Env)$ more than the (weighted) sum of its parts.

We could now assume an extreme case of relative fitness in which each environmental state has only one surviving type with superior fitness, $w(\{g_1 = fittest\}|e) > 0$, while in this environmental state all other types die out with $w(\{g_1 = unfit\}|e) = 0$. Only one type fits the particular environment. In this case the fittest type represents the entire population after updating, $p(\{g_1^{t+1} = fittest\}|e) = 1$, and equation (8.2) simplifies to:

$$\log w(\overline{g_0}) = E^{t+1,e}[\log w(\overline{g_1}, e)] - \sum_e 1 * p(e) * \log\left(\frac{p(e)*1}{p(e)*p(\overline{g_1}^t|e)}\right)$$

$$= E^{t+1,e}[\log w(\overline{g_1}, e)] + \sum_e p(e)\log p(e) - \sum_e p(e)\log\frac{p(e)}{p(\overline{g_1}^t|e)}$$

$$= E^{t+1,e}[\log w(\overline{g_1}, e)] - H(Env) - D_{KL}\left(P(e)\|P(\overline{g_1}^t|e)\right) \qquad (8.6)$$

Equation (8.6) is a well-known result derived by Kelly in 1956 *(39)*. Since both H and D_{KL} are always positive, it says that the attainable average population fitness is reduced by the level of uncertainty inherent in the environment, $H(Env)$, as well as by the distance of the distribution of the evolving population from its environment, D_{KL}. Among other things, this leads to the celebrated result in portfolio theory that a proportional bet hedging strategy maximizes population fitness (with $P(e) = P(\overline{g_1}^t|e)$ and therefore $D_{KL} = 0$, for related work see also *(5-7)*). As such, equations (8.3) and (8.4) are generalizations of Kelly's special case result *(39)*, which also holds for the less restrictive case with variety of growth factors among types (for more recent results on Kelly's criteria see *(5-7,40-43)*).

It is important to notice that a multilevel decomposition of fitness on a lower level implies that we identify lower level types with distinguishable fitness. This also implies the existence of distinct lower environmental states within the higher level environmental states that interact with this more fine-grained level. For example, while all kinds of flowering plants thrive in a sunny environment, Fig. 3 supposes that different kinds of flowering plants thrive more in sunny days that are windy, while others have superior relative fitness on sunny days that are burning hot. If more fine-grained types have distinguishable relative fitness within the coarse-grained environmental state, a more fine-grained distinction of the environment becomes a requisite in order to calculate the conditional fitness values on the next lower level. The number of

distinguishable environmental states that affect evolutionary dynamics grows with the number of

relative fitness patterns that are identifiable from more fine-grained distinctions of types. This

implies that the number of distinguishable environmental states conditioned on the superiority of

a specific type, or—the other way around—distinguishable types conditioned on an

environmental state, grows with more detailed levels of fine-graining:

$$|(e_l|g_l)| < |(e_{l+1}|g_{l+1})| \qquad (9.1)$$

$$|(g_l|e_l)| < |(g_{l+1}|e_{l+1})| \qquad (9.2)$$

This leads to a fresh look at Ashby's law of requisite variety *(44)*. The level of fine-graining of

distinguishable environmental states in time, and the number of distinguishable types of an

evolving population in space go hand in hand. The exploitation of environmental variety requires

type variety, and type variety implies that there must be discernable environmental variety. This

becomes very clear in the extreme case of Kelly's optimized bet-hedging strategy. In this

extreme case, there is only one optimized type for each environmental state, leading to a one-to-

one relation between the number of types and discernable environmental states. The argument

presented here generalizes this logic of space-time fine-graining (for an illustrative example, see

Fig. 3; for an empirical example see (S.2.7)).

Discussion

Until now, evolutionary dynamics have mainly been analyzed with the help of differential

equations *(27)* or in terms of variances and covariances *(33,34)*. Why the need to reformulate

evolutionary dynamics in terms of information theory? For one, it formalizes enduring notions

that link evolution, informational complexity, and ideas like 'negentropy' *(1,25)*. We derived an

intuitive definition of 'fit-ness' as an informational 'fit' between the evolving population and its

environment. This provides new meaning to 'fit-ness'. Besides, it has long been suggested that "discoveries might be explained as the outcome of the process of switching the problem representation to a different ontological category" *(45; p. 430; also 46,47)*. In this initial exploration we started by reformulating longstanding concepts of evolutionary theory in terms of information theory, including Fisher's fundamental theorem of natural selection, the Price equation, the role of the reversed replicator equation, multilevel emergence, and evolutionary sustainability. This led to new conceptualizations of well-known phenomena, such as the definition of change in fitness in a stable setting as the sum of descriptive complexities of the present from the perspectives of the past and future (equation (5)), and led to the generalization of Kelly's bet-hedging criterion, and a space-time version of Ashby's law of requisite variety. These insightful results underline the urgency to take a fresh look at all kinds of evolutionary dynamics.

One ongoing discussion in which information theoretic approaches to evolution have much potential is the the question about the adequate carrier of information (i.e. the adequate levels of selection *(48-50)*). The variables chosen to define levels and their types influence the resulting constellations of fitness. Information theory is an adequate language to analyze the choice of variables and the amount of information they contain. For example, Fig. 1A classified the evolving population according to variables defined by the United Nations Standard International Trade Classification (SITC, rev.2), but who says that evolution cares about this kind of classification? What exactly are the levels and groups of selection that process information? The presented derivations provide a foundation to tackle such and other questions with the help of a new set of well-understood analytical tools.

References and Notes:

1. Smith, JM. & Szathmary, E (1997) *The Major Transitions in Evolution*. (Oxford University Press).

2. Frank, SA (2009) Natural selection maximizes Fisher information. *Journal of Evolutionary Biology* **22,** 231–244.

3. Frank, SA (2012) Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *Journal of evolutionary biology* **25,** 2377–96.

4. Adami, C (2012) The use of information theory in evolutionary biology. *Annals of the New York Academy of Sciences* **1256,** 49–65.

5. Donaldson-Matasci, MC., Bergstrom, CT & Lachmann, M (2010) The fitness value of information. *Oikos* **119,** 219–230.

6. Rivoire, O & Leibler, S (2011) The Value of Information for Populations in Varying Environments. *Journal of Statistical Physics* **142,** 1124–1166.

7. Permuter, H-H., Kim, YH & Weissman, T (2011) Interpretations of Directed Information in Portfolio Theory, Data Compression, and Hypothesis Testing. *IEEE Transactions on Information Theory* **57,** 3248 –3259.

8. Krakauer, DC (2011) Darwinian demons, evolutionary complexity, and information maximization. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **21,** 037110–037110–12.

9. Fisher, RA (1930) *The Genetical Theory of Natural Selection*. (Clarendon Press, Oxford).

10. Price, GR (1972) Fisher's 'fundamental theorem' made clear. *Ann. Hum. Genet.* **36,** 129–140.

11. Frank, SA (1997) The Price Equation, Fisher's Fundamental Theorem, Kin Selection, and Causal Analysis. *Evolution* **51**, 1712–1729.

12. Edwards, AWF (2000) Fisher information and the fundamental theorem of natural selection. *Rendiconti Scienze B* **134,** 1–15.

13. Fisher, RA (1925) Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* **22,** 700–725.

14. Frieden, BR, Plastino, A & Soffer, BH (2001) Population Genetics from an Information Perspective. *Journal of Theoretical Biology* **208,** 49–64.

15. Frank, SA (2012) Natural selection. III. Selection versus transmission and the levels of selection. Journal of Evolutionary Biology **25**, 227–243.

16. Shannon, CA (1948) Mathematical Theory of Communication. *Bell System Technical Journal* **27,** 379–423, 623–656.

17. Kullback, S & Leibler, RA (1951) On Information and Sufficiency. *Ann. Math. Statist.* **22,** 79–86.

18. Cover, TM. & Thomas, JA (2006) *Elements of Information Theory*. (Wiley-Interscience).

19. Clausius, R (1865) Über verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechinischen Wärmetheorie. *Annalen der Physik und Chemie* **CXXV**.

20. Boltzmann, L (1872) Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte Akademie der Wissenschaften* **66**, 275–370.

21. Zurek, WH (1990) *Complexity, Entropy and the Physics of Information*. (Westview Press).

22. Gibbs, JWA (1873) Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces. *Transactions of the Connecticut Academy* II, 382–404.

23. Jaynes, ET (1957) Information Theory and Statistical Mechanics. *Phys. Rev.* **106**, 620.

24. Crow, JF. & Kimura, M (1970) *An introduction to population genetics theory*. (Blackburn Press).

25 Schrödinger, E (1944) *What is life? Mind and Matter*. (Cambridge University Press).

26. Brillouin, L. (1953) The Negentropy Principle of Information. *Journal of Applied Physics* **24**, 1152–1163.

27. Nowak, M. A. (2006) Evolutionary Dynamics: Exploring the Equations of Life. (Belknap Press of Harvard University Press).

28. Jensen, J. L. W. V. (1906) Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.* **30**, 175–193.

29. Garrity et al. Draft BioCode (2011) Principles and Rules Regulating the Naming of Organisms. *Bulletin of Zoological Nomenclature* **68**, 10–28.

30. UNSD (United Nations Statistics Division) (2012) *UN COMTRADE database*. (United Nations Statistics Division).

31.Baum, JAC & McKelvey, B (1999) *Variations in Organization Science: In Honor of Donald T Campbell*. (Sage Publications, Inc,).

32.Aldrich, H & Ruef, M (2006) *Organizations Evolving*. (Sage Publications Ltd).

33. Price, GR (1970) Selection and Covariance. *Nature* **227,** 520–521.

34. Frank, SA (2012) Natural selection. IV. The Price equation. *Journal of Evolutionary Biology* **25**, 1002–1019.

35. For lack of common notation for multilevel nesting, after discussions with colleagues I chose the notation [...\...], since nesting is related to but not necessarily and not for all cases identical to both conditioning [...|...] and division [.../...]. In our purposes of conditional relative entropy in equation (4.1), however, multilevel nesting behaves equal to traditional conditional probabilities (see also Supporting Material (S.1)).

36. $J(P^t, P^{t+1})$ also approaches Fisher information *(13)* in the limit of small changes in the underlying parameter *(3)*.

37. Hilbert, M (2013) Linking Information, Knowledge and Evolutionary Growth: A multilevel interplay between natural selection and informed intervention. *Presented at 'Advances in Dynamic Economics and Development: Economic and Complexity'*.

38. Bateson, G (1972) *Steps to an Ecology of Mind*. (Random House, New York); p. 459.

39. Kelly, J (1956) A new interpretation of information rate. *Bell System Technical Journal* **35**, 917–926.

40. Kussell, E & Leibler, S (2005) Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. *Science* **309,** 2075–2078.

41. Donaldson-Matasci, M, Lachmann, M & Bergstrom, CT (2008) Phenotypic diversity as an adaptation to environmental uncertainty. *Evolutionary Ecology Research* **10,** 493–515.

42. Cherkashin, D, Farmer, JD & Lloyd, S (2009) The reality game. *Journal of Economic Dynamics & Control* **33,** 1091–1105.

43. Rainey, PB *et al.* (2011) The evolutionary emergence of stochastic phenotype switching in bacteria. *Microb. Cell Fact.* **10,** S14.

44. Ashby, WR (1956) *An introduction to cybernetics*. (Chapman & Hall). at <http://pcp.vub.ac.be/books/Introcyb.pdf>

45. Chi, M & Hausmann, R (2003) in *The International Handbook on Innovation* (Shavinina, L. V.) 430–444 (Elsevier).

46. Polya, G (2008) *How to Solve It: A New Aspect of Mathematical Method*. (Princeton University Press).

47. Tao, T (2006) *Solving Mathematical Problems : A Personal Perspective: A Personal Perspective*. (Oxford University Press).

48. Keller, L (1999 *Levels of selection in evolution*. (Princeton University Press).

49. Wilson, RA (2007) in *Philosophy of biology* (Matthen, M. & Stephens, C.) 141–162 (Elsevier).

50. Hilbert, M, Oh, P & Monge, P (2013) Evolution of what? A network approach to the definition of evolving populations. *XXXIII Sunbelt Conference of INSNA*.

51. Feenstra, RC, Lipsey, RE, Deng, H, Ma, AC & Mo, H (2005) *World Trade Flows: 1962-2000*. (National Bureau of Economic Research).

Fig. 1. Fitness is incrementally decomposed into relative information over multiple levels. **(A),** Equation (3.2) decomposition of log_2 fitness of international export volume of 512 Canadian products over 5 levels according to the official multilevel Standard International Trade Classification (SITC) rev.2 of the United Nations between 1962 and 2000 in US\$ exported *(30,51)*. $E_i[...]$ stands for Expected value. **(B),** Time averages of multilevel D_{KL} terms of equations (3.1) (striped column) and (3.2) (filled column) of number of YouTube views of 56 online video clips of the nonprofit news channel DemocracyNow! over 6 binary-split levels, classified in types according to the number of comment per view (an indicator of the level of controversy of a clip). Presented as time averages of D_{KL} over 24 collected periods. Round inserts are zoomed-in excerpts of the same respective columns.
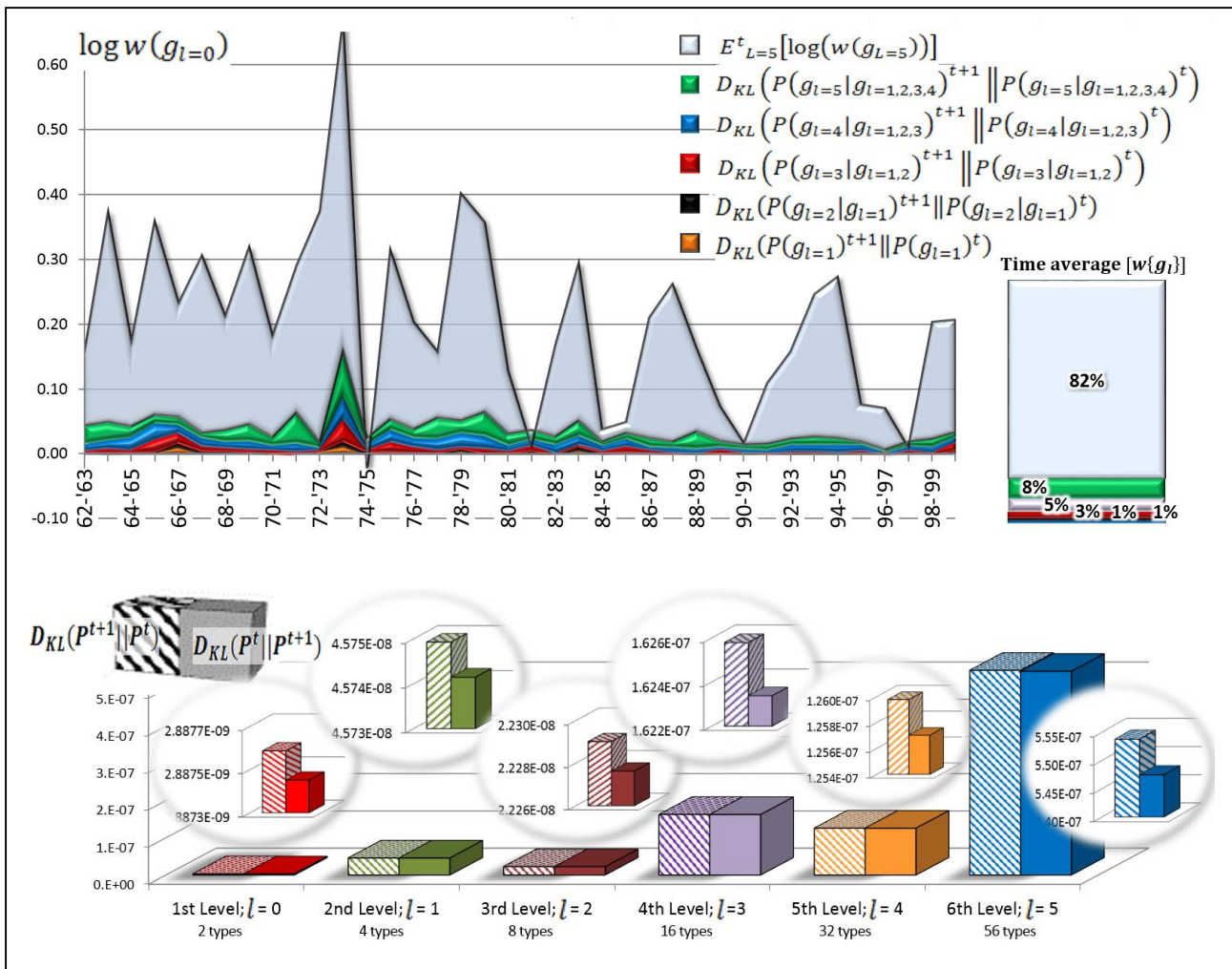
Fig. 2. Equation (7.2) decomposition of log fitness of fundraising amounts of 120 entrepreneurs from the international Microfunds platform Kiva.org (in US\$ fundraised over a 4 day period) fine-grained in 7 binary-split levels classified in types according to the size of the intermediating partner organization (in total loans collected). The left column presents the fundraising effort from 08/07/13 to 08/09/13, and the right column from 08/09/13 to 08/11/13. Left axis: $[\log w(g_0{}^{t,t+1})]$; Right axis: multilevel expression of $\left[D_{KL}\left(P_l{}^{t+1} \| P_l{}^{t+2}\right) - D_{KL}\left(P_l{}^{t+1} \| P_l{}^{t}\right)\right]$. While growth rates slow down during the second period, the change of the evolutionary population constellation (measured in ΔD_{KL}) slows down even more drastically.
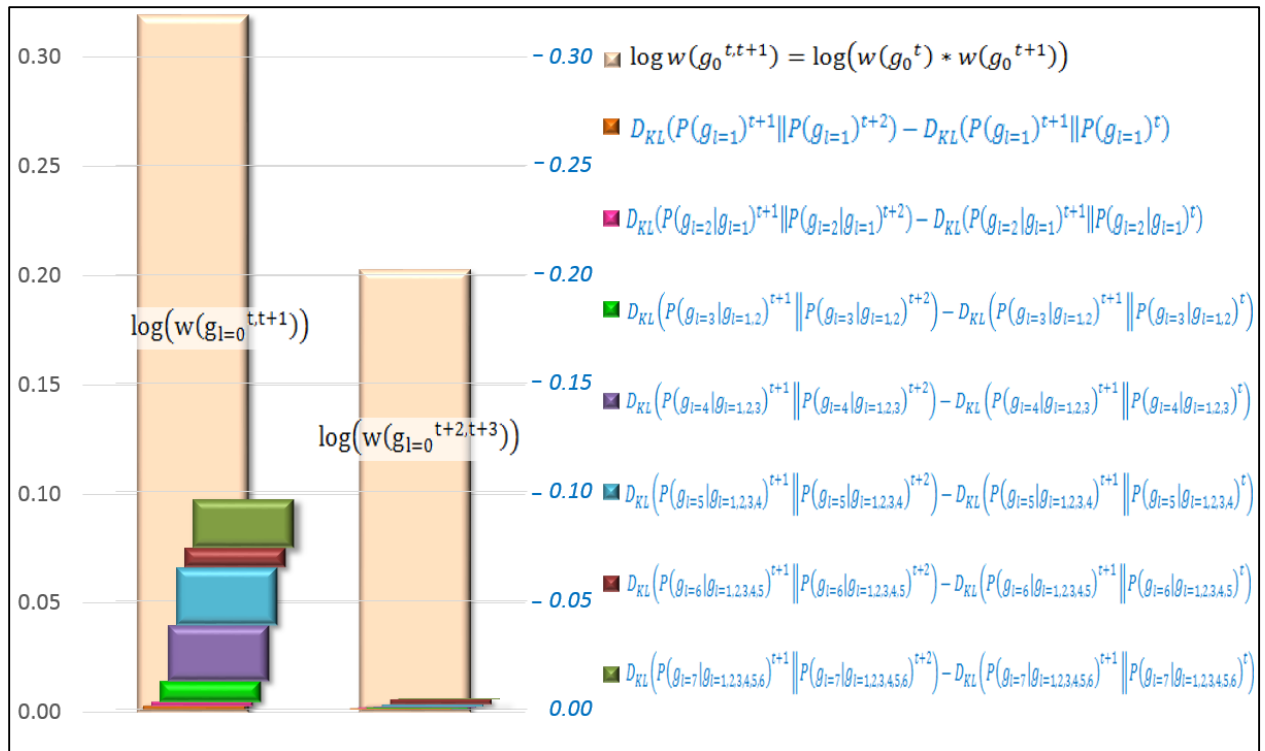
Fig. 3. Illustrative schematization of space-time requisite variety between the evolving system and its environment. **(A),** on the coarse-grained level 1 the figure distinguished between $p(g_{l=1} = {}'A'ngiosperms)$ and $p(g_{l=1} = {}'C'onifers)$, whose fitness responds differentially to the absence and presence of sun. It is assumed that conifers have higher relative fitness in the absence of sun, and vice versa. The environment is assumed to be distributed according to $p(e = \overline{sun}) = 0.4$, and $p(e = sun) = 0.6$. Empirically, the distinguishable relative fitness is what calls attention to the fact that the presence of the sun represents a "difference that makes a difference" *(28).* **(B)**, Level 2 fine-grains the population into two different kinds of flowering plants, and two different kinds of conifer. The illustrative examples assumes that in the absence of sun, the tall skinny conifer d has superior fitness to its counterpart conifer c, while the relative fitness of flowering plants happens to depend on a further level of environmental fine-graining: the presence of snow or rain. $w(g_{l=2}|e = snow, \overline{sun}) \neq w(g_{l=2}|e = rain, \overline{sun})$. As such, the detection of changing relative fitness patterns on a more fine-grained level of the evolving population, lead to fine-graining in the environmental states. In the presence of sun, the relative fitness constellation of distinct flowering plants depend on the intensity of the sun, while different conifers thrive in the absence and presence of wind. As a more formal proof of concept of these illustrative notions, Supporting Material (S7) applies this logic to the empirical case of the evolution of USA exports of goods in US$ grouped on the first level into manufacturing and non-manufacturing good, and on the second level into manufactured goods and machinery (both manufacturing), and edibles and substances (both non-manufacturing).
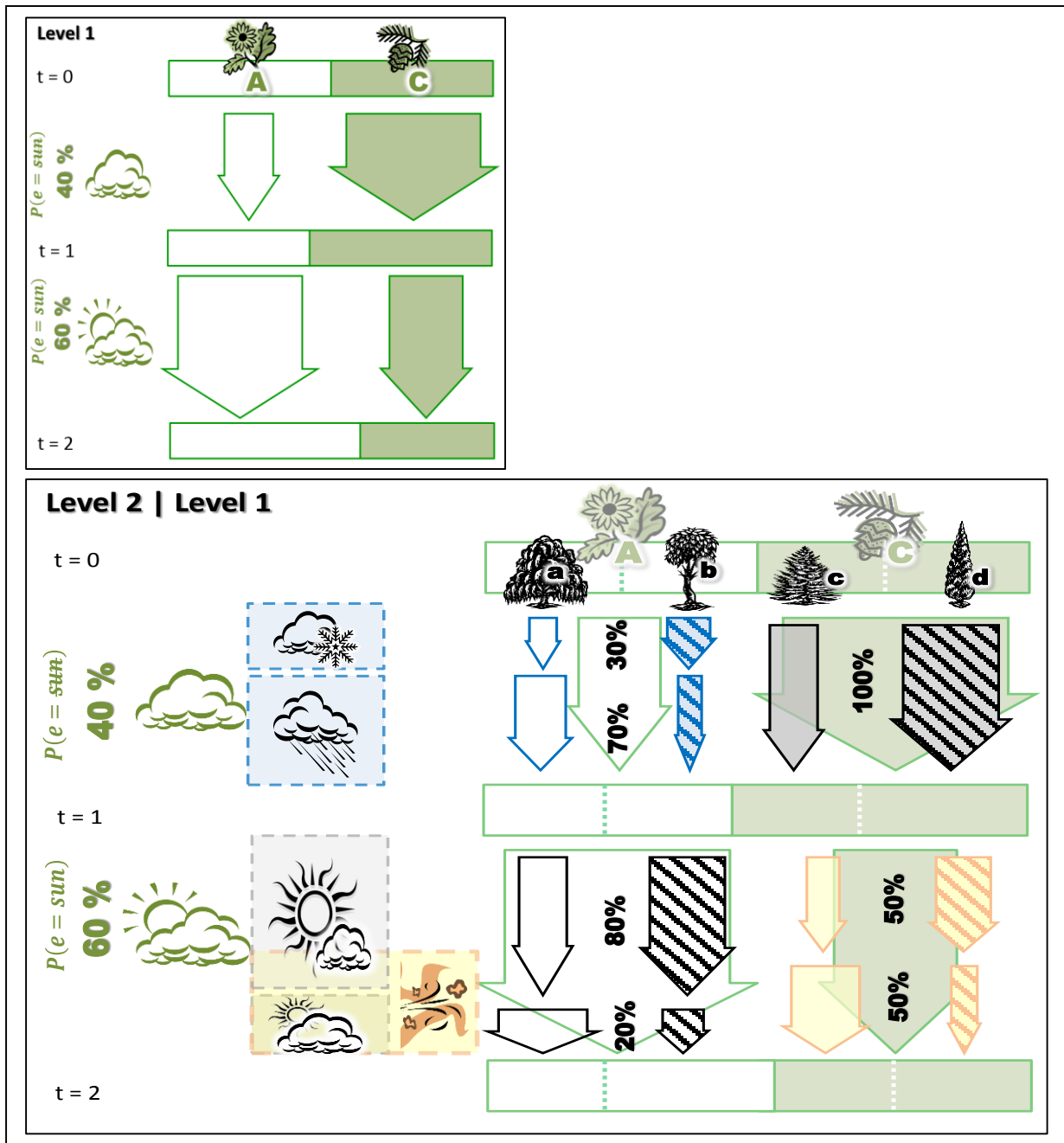
Supplementary Information:
Sections: (S1) – (S4)
Figures: Fig. SF.1 – Fig. SF.6
Equations: (S.2.1.1) – (S.2.1.7)
Supplementary Data 1 & 2

34