### Title

An Information Theoretic Decomposition of Fitness: Engineering the Communication Channels of Nature and Society




### Author

Hilbert, Martin



### Publication Date

2015-03-15

Peer reviewed

# An information theoretic decomposition of fitness:

# engineering the communication channels of nature and society

March 2015

Martin Hilbert

University of California, Davis, CA, 95616, USA; hilbert@ucdavis.edu

**ABSTRACT**

**Formal analyses in biology, ecology and economics typically bypass the intuitive interpretation of 'fit-ness' as environmental 'fit' and hastily equate fitness with the resulting rate of reproduction, economic pay-off, etc.. But evolutionary growth also has a direct interpretation as environmental 'fit', as it can be expressed in terms of Shannon's mutual information. This presents evolutionary change as a communication channel between the population and its environment. Optimal growth quantifies the amount of structure in the updated population that unequivocally comes from the environment. Turning this finding around, fitness can be optimized by searching for non-confusable signals between both. Full channel capacity is achieved with specialized types for each environmental state (noiseless channel). Just like technological communication channels thrive on source knowledge, fitness can be increased with environmental knowledge. This establishes a formal link between evolution in biological and social populations, and long-standing engineering efforts in the optimization of communication channels.**

Martin Hilbert; hilbert@ucdavis.edu

"What do we mean when we say that A communicates with B? We mean that the … acts of A have induced a … state in B"[1]. Evolution can be understood as communication between the environment and the evolving population. The environment induces structure in the population through the process of evolutionary change. The better the 'fit' between the environment and the evolving population, the higher the 'fit-ness'. The result is the growth of the population that exploits this fit. Going back to the initial "mathematical theory of natural and artificial selection"[2] from 1924, it is this resulting growth factor (or its normalized equivalent) that is traditionally taken to represent 'fitness' in the mathematical models of ecology, social evolution, and economic evolution (rate of reproduction, game pay-off, economic returns, etc.)[3,4,5]. It turns out that evolutionary growth can also directly be equated with one of the most fundamental metrics of 'fit', which is informational fit. This can be formalized with information theory[1,,6,7]. Information theory is a branch of mathematics that goes back to Shannon's seminal work[6] from 1948. It provides formal concepts and metrics to deal with fundamental questions of communication processes, such the question about the ultimate limit of data compression (the elimination of redundant data to obtain pure information), and the ultimate transmission rate (the channel capacity). Recent studies have shown that aspects and parts of the evolutionary process can be expressed in information theoretic terms[8,9,10,11,12,13,14,15,16,17]. Information theory seems to be useful to explore evolutionary dynamics. Existing contributions work with partial aspects of evolutionary dynamics (such as natural selection)[8,9,16]; adopt limiting modeling assumptions (such as a diagonal fitness matrix) [13-15]; or work with a mix of analytical tools and metrics from economic decision theory, matrix algebra, and information theory[10,11] that provide a picture that is both incomplete and somewhat involved, but at the same time very promising. The following shows how the complete expression of evolutionary fitness can be represented purely in information theoretic terms. The result is a decomposition of fitness' growth factor using information metrics. The interpretation turns out to be quite intuitive and provides a natural interpretation of the concept of evolutionary 'fit-ness' as informational 'fit'.

# Results

## Basic multilevel decomposition.

Following existing conventions, we define fitness' factor of reproduction as $w = \frac{units\ at\ time\ t+1}{units\ at\ time\ t}$, whereas the units of the evolving population can represent number of parents and offspring, game bets and payoff in $, portfolio investments and returns, etc. The different population types and environmental states are represented by random variables with distributions $P(G)$ and $P(E)$, respectively. Their interaction can be represented by a traditional fitness matrix[18], such as the one illustrated in Figure 3b. The updating of the population due to evolutionary change in a specific environment is given by the so-called replicator equation[4,16]: $p'(g|e) = p(g|e)\frac{w(g,e)}{\bar{W}(e)}$, whereas $p(g|e)$ refers to the share of type $g$ conditioned on the environmental state $e$; the apostrophe $'$ indicates the updated generation after reproduction (a commonly used short-hand for the long-hand notation $t+1$); $w(g,e)$ refers to the fitness of a specific type in a specific environmental state; and the single overbar represents the average over all types in a specific environment: $\bar{W}(e) = \sum_g p(g|e) * w(g,e)$. The long-term average population fitness in space and time, $\bar{\bar{W}}$, is the weighted geometric mean of the population fitness over all environmental states: $\bar{\bar{W}} = \prod_e \bar{W}(e)^{p(e)}$.

As outlined in Methods, we can represent this long-term average fitness as:

$$\bar{\bar{W}} = \prod_{g,e} w(g,e)^{p'} * 2^{-D_{KL}(P'(g,e)\|P(g,e))} \qquad (1)$$

$$\log \bar{\bar{W}} = E_{p'}[\log w(g,e)] - D_{KL}(P'(g,e)\|P(g,e)) \qquad (2)$$

The only difference between both equations is that the logarithm is taken in equation $(2)$. This represents growth factors on a logarithmic scale (what is traditionally referred to as Malthusian fitness[19,20]). The logarithms of base 2 gives an intuitive information-theoretic interpretation in bits and represents fitness in terms of the number of population doublings at each time step. The structure of the equations is two-fold and involves some kind of aggregate of type fitness (a product in equation $(1)$, which becomes an

expected value in equation (2)), and an information theoretic second term $D_{KL}$. $D_{KL}$ is the relative entropy, or Kullback-Leibler divergence, one of the most fundamental metrics of information theory[21]. It is an unsymmetrical measure of divergence, in this case between the distribution of the original and the updated population. In our case, $D_{KL}$ quantifies the divergence that occurs during updating. This agrees with the interpretation of recent research that shows that $D_{KL}$ that it is logarithmically related to the variance in fitness, and as such to Fisher's fundamental theorem of natural selection[19] (a traditional way to quantify the 'speed/magnitude' of evolutionary change during updating)[8,9].

A formal interpretation of $D_{KL}$ stems from large deviation theory[Error! Bookmark not defined.]. $D_{KL}$ quantifies the probability that a population ends up with an updated population $P'$ when originating from an original population $P$. Imagine an ecosystem with two types of finches (long and short peaked) represented by the random variable $G$, which are distributed according to $P(G)$, with $p(g = long) = 0.4$ and $p(g = short) = 0.6$. The law of large numbers tell us that with many observations the probability of observing $40\ \%$ long-peaked finches in this ecosystem is on average close to 1. The probabilistic version of the law of large numbers (the so-called asymptotic equipartition property) tells us that with $n$ observations, there are about $2^{nH(G)}$ such sequences, each with probability $\frac{1}{2^{nH(G)}}$, whereas $H$ is Shannon's absolute entropy, a metric that quantifies the uncertainty of the distribution[6]. What is the probability that we would observe (for example) $20\ \%$ of long-peaked and $80\ \%$ of short-peaked finches? This can also occur, even so it is exponentially unlikely (since such sequences are outside of the typical set). It would however be very likely if the real distribution were $P'(G)$, with $p'(g = long) = 0.2$ and $p'(g = short) = 0.8$. The probability that such largely deviated observations occur therefore depends on the distance between both distributions, which is quantified by the Kullback-Leibler relative entropy. Sanov's theorem[22] tells us that this occurs on average $2^{nD_{KL}(P'(g)\|P(g))}$ number of times, each with probability $2^{-nD_{KL}(P'(g)\|P(g))}$, whereas $P$ is the true distribution, while $P'$ is the observed distribution[Error! Bookmark not defined.].

Note that the setup of both equations is recursive, as the same decomposition that is applied to the average higher population level fitness on the left-hand side $W$, can be applied to the lower level fitness of specific types on the right hand side $w(g,e)$. This would imply to decompose the fitness of each specific type into more fine-grained subtypes on a lower taxonomic level. In this case the chain rule for relative entropy[Error! Bookmark not defined.] sums the conditional $D_{KL}$ terms of each level to the relative entropy on the lowest registered level of fine-graining. Also note that the average logarithmic population fitness can be completely decomposed into relative entropy $D_{KL}$ if the weighted geometric mean fitness of the updated types is equal to 1: $\prod_{g,e} w(g,e)^{p'(g,e)} = 1$.

## A generalization of Kelly's result from 1956.

Equation (2) is a generalization of Kelly[23] interpretation of information rate from 1956. Kelly's criteria criteria shows the long-term superiority of bet-hedging strategies (also known as stochastic switching) and has become an important building block in portfolio theory, evolutionary finance and ecological thinking [10,11,13-15,24,25,26,27,28,29]. However, its setup is restricted to a diagonal fitness matrix. Only the fittest type survives in each environment, growing with diagonal fitness values ${}^{d}W$. From the descriptive approach of the previous equations, this implies that only one type exists after updating ($p'(g = fittest, e) = 1$), and equation (2) simplifies Kelly's result[30]:

$$\log \bar{\bar{W}} = E_e[\log {}^{d}W] - H(e,g) = E_e[\log w] - H(E) - D_{KL}(P(e)||P(g)) \qquad (3)$$

$H(e,g)$ is the cross-entropy between the environment and the population. It consists of the absolute entropy of the environment $H(e)$, and the divergence between the distribution of the environment and the original population before updating. Kelly used this result to show that in his setup with a diagonal fitness matrix, the optimal growth factor can be achieved through a proportional bet-hedging strategy that assures that the distribution of the population exactly matches the environmental distribution $P(E) = P(G)$. This is the only case for which the non-negative metric $D_{KL}(P(e)||P(g))$ becomes zero.

This way, the achievable fitness in equation (3) is only compromised by the inherent uncertainty of the environment $H(E)$, which is also a non-negative metric. If the entire dynamic of the future environment is known, the highest achievable growth rate consists of the expected value of the diagonal fitness values per environment $\log \overline{\overline{W}} = E_e[\log {}^{d}W]$.[30]

## Optimal Fitness.

Equations (1) and (2) are generalizations of Kelly's result, because they do not require a diagonal fitness matrix. However, they are merely descriptive of the updating process and, and unlike Kelly's result, not say anything normative about the optimality of updating. We now ask about the optimal population fitness in mixed fitness landscapes with more than one non-fatal type per environmental state (thereafter referred to as mixed fitness landscapes, in contrast to Kelly's diagonal fitness matrix setup). It turns out that optimality is achieved in mixed fitness landscapes if the joint average population before updating is equal to the independent state between updated population and environment, $p(g,e) = p'(g) * p(e)$. In this case, $D_{KL}$ from equations (1) and (2) turns into Shannon's much celebrated metric of mutual information[1,6,7] (see Methods).

$$\log \overline{\overline{W}} = E_{p'}[\log w(g,e)] - I(G';E) \qquad (4)$$

The mutual information $I(G';E)$ measures the shared information between the environment and the average updated population. It quantifies the amount of structure in the updated population $G'$ that is assured to come from the environment $E$. This is because of the nature of joint typicality of both sets (see Fig. 1). The likelihood of finding a non-confusable, non-overlapping input signal from the environment in the updated population is the conditional uncertainty of the updated population given the environment, divided by the uncertainty of the population: $\frac{2^{nH(G'|E)}}{2^{nH(G')}} = 2^{-nI(G';E)}$ (Fig. 1). Because mutual information is symmetric, one can equivalently say that the probability that a randomly chosen environmental signal is jointly typical with the updated future generation of the evolving population is about $\frac{2^{nH(E|G')}}{2^{nH(E)}} = 2^{-nI(G';E)}$

(the uncertainty of the environment, given an updated population type, divided by the uncertainty of the environment). The chance of capturing such signal is small, but might be found through purely blind (neo-Darwinian) evolution through random mutations.

The outline described in Methods shows that it is always true that growth optimization implies the condition $D_{KL} = I$. However, the revers requires that the types and environmental states are defined in a way that makes them linearly independent. This seems to be a reasonable demand, as redundant types or environmental states should be merged.

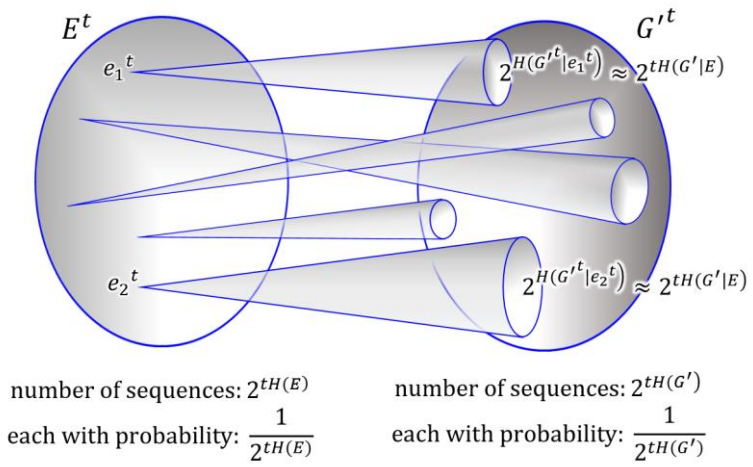


**Figure 1 | Channel after t uses.** The left presents the typical set of the environmental states $E$ and the right the typical set of the average updated future generation $G'$, both over a large number of periods $t$. The transmission over a noisy channel induces uncertainty to the identification of each environmental state during reception by the population. For example, the uncertainty that the environmental state $e_1$ is sent over the channel is the conditional entropy of $G'$ given that $e_1$ was sent: $H(G'^t|e_1{}^t)$. According to the asymptotic equipartition property, there are approximately $2^{H(G'^t|e_1{}^t)}$ of those. The total number of typical $G'$ sequences is $\approx 2^{tH(G')}$. Restricting ourselves to the subset of channel input such that the corresponding typical output sets do not overlap, we can bound the number of non-confusable inputs by dividing the size of the typical output set by the size of each typical-output-given-typical-input set: $2^{tH(G'|E)}$. The result shows that the total number of disjoint and non-confusable sets is less than or equal to: $2^{t(H(G')-H(G'|E))} = 2^{t\,I(G';E)}$. For a detailed proof see any standard textbook on information theory, such as references 1 or 7.

This shows that optimal growth is achieved by assuring that the structure in the updated population $G'$ stems from unequivocal signals of the environment $E$. The setup as a communication channel in Fig.1 also

gives a natural assignment of which side is the environment and which side the affected population. The environment is the sender, and the population the receiver. As in any communication process, roles may be switched (which will switch the conditionality of entropies in our interpretation).

## The more you know, the more you can grow.

A complementary interpretation of the foregoing equations can be obtained by asking how the information theoretic decomposition changes if we obtain additional side information. The two possible sources of additional side information are observations of the past that influence current and future dynamics ('memory') and observations of third events that correlate with current and future dynamics ('cues'). This introduces a new conditioning variable $C$. Conditioned on the realization of this side information, the joined distributions can change and we end up with fine-tuned strategies for each conditioned case. It is a fundamental theorem in information theory that conditioning reduces entropy[1]. In Kelly's setup it reduces environmental uncertainty through $H(E) \geq H(E|C)$, and therefore increases the achievable fitness in equation $(3)$. The increase is equal to the mutual information between the cue and the environment: $H(E) - H(E|C) = I(E;C)$ (which has been termed the 'fitness value of information' [11,15,23]). This can also be visually represented as the overlapping intersection in the form of the Venn diagram shown in Fig. 2a (also called I-diagrams[1,7,31,32]).

Our generalized case of Kelly's result reveals an important third variable in this interaction: the average updated population $G'$. The importance of this variable arises for mixed fitness landscapes. As shown in Fig. 2, this is not obvious in Kelly's special case, because in this case $H(E) = H(G')$. However, in the case of optimal growth in mixed fitness landscapes it turns out that the three variables form a Markov chain $E \leftrightarrow G' \leftrightarrow C$, whereas the cue and the environment are conditionally independent given the average updated population (see Fig. 2b). In information theoretic terms, this means that there is no mutual

information between the cue and the environment given the updated population: $I(E;C\ |G') = 0$. In other words, optimal growth implies that all structure is absorbed by the updating during optimal growth. This leads to a conditional version of equation $(4)$:

$$\log \overline{\overline{W}}_{|c} = E_{p'(g,e\,|c)}[\log w(g,e\,|c)] - I(G';E\ |C) \qquad (5)$$

The fitness value of the cue is obtained by the difference between the fitness without cue (equation $(4)$) and with cue (equation $(5)$). The expected value term cancels and we obtain the three-way mutual information between all three variables: $\log \overline{\overline{W}}_{|c} - \log \overline{\overline{W}} = -I(G';E\ |C) + I(G';E) = I(G';E;C)$. It is important to notice that three-way information can in principle be negative[1,7,31,32]. However, in our case Markovity assures that it is positive, since $I(G';E;C) = I(E;C)$ and two-way mutual information is always positive (this is visualized by Fig. 2 and can formally be shown with the data processing inequality[1,7]). Therefore, any cue must increase fitness on average. By how much? By the mutual information between the environment, the average updated population, and the cue. This three-way mutual information quantifies the generalized fitness value of information.

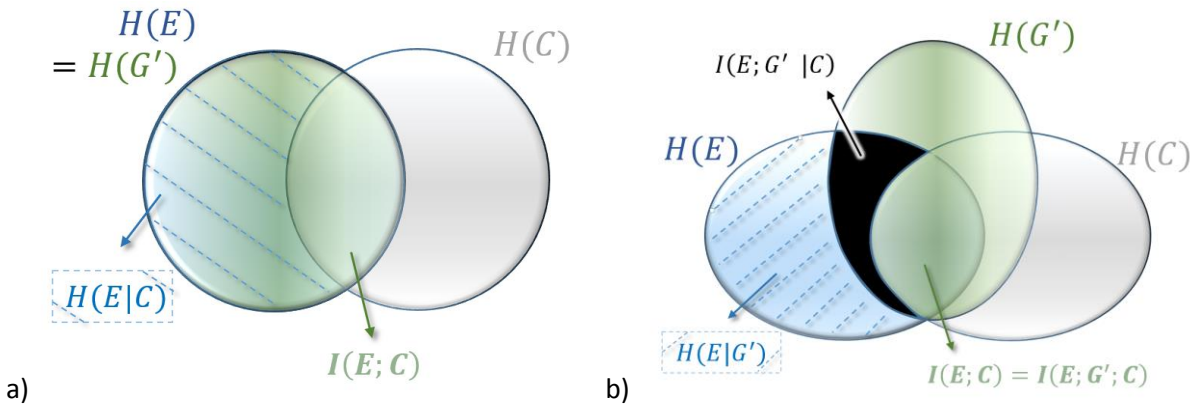


**Figure 2 | Venn diagram / I-diagram representation of mutual information.** Mutual information is defined as any intersection of the represented entropies. **(a)** In the two-variable case, it can be calculated as $H(E) - H(E|C) = I(E;C)$. For example, the environmental random variable $E$ could have the realizations sun and rain, while the side information $C$ consists of past patterns of sun and rain, or of correlated cues like wind. $H(E)$ is the uncertainty of the random variable sun/rain, while $H(E|C)$ could be the uncertainty of sun/rain conditioned on the presence and absence of wind. The mutual information is the difference between both uncertainties. It is always positive in the two-variable case, as conditioning reduces uncertainty. **(b)** A similar logic holds for the three variable case. One way to calculate the joint intersection of all three variables would be: $I(G';E;C) = H(E) - H(E|G') - I(G';E\ |C)$. In our case, the three involved variables form a Markov chain $E \leftrightarrow G' \leftrightarrow C$. This implies that $E$ and $C$ do not have any mutual information outside of $G'$ ($G'$ absorbs all common structure through optimal growth), or $I(E;C\ |G') = 0$. This can be shown by the reformulation $I(E;C\ |G') = H(E|G') - H(E\ |C,G')$ (which holds in general). In our case $H(E|G') = H(E\ |C,G')$, since the addition of some cue does not affect the distribution of environmental states (the mere observation of wind does not change an existing sun/rain pattern).

## The complete decomposition with constraints.

Recent research has proposed to extend Kelly's result for diagonal fitness matrices to more general mixed fitness landscapes by introducing a "hypothetical weighting matrix"[10,11]. This weighting matrix represents any non-fatal fitness as a combination of optimized fitness values from a hypothetical diagonal fitness matrix that only contains one optimal type per environment while all others are fatal: ${}^{d}_{hyp}W$. In other words, it assumes a hypothetical world with one perfectly specialized type per environment and proposes that any existing type is a combination of those pure states. This way, any mixed fitness landscape can be represented as a weighted mix of a purely diagonal matrix. This proposes to rewrite the long-term growth

rate of type $g_i$ as: $\prod_e\{w(g_i,e)\}^{p(e)} = \prod_e\{\sum_e p(e|m_i) * {}_{hyp}^{d}W(e)\}^{p(e)}$, whereas the "hypothetical weighting matrix" $P(E|M)$ is a stochastic matrix that weighs each type over all different environmental states in such a way that achieves this theorized equivalence.

Using this approach, Donaldson-Matasci et al.[10,11] have found that proportional bet-hedging also works for mixed fitness landscapes. However, the proportionality is not in one-to-one correspondence with the environmental distribution (but has to adjust for the distortion of a mixed fitness landscape). They also found that it only works within a certain range of population constellation. They termed this region the "region of bet-hedging". Outside this region the achievable optimal growth rate is diminished by the constraints of the existing mixed fitness landscape.

This important insight can naturally be incorporated into our generalized framework. The arising constraints can be expressed purely in information theoretic terms (see Methods).

$$\log \overline{\overline{W}} = E_e\left[\log {}_{hyp}^{d}W\right] - H(E|G') - D_{KL}(P'(e|g)\|P(e|m)) - D_{KL}(P'(g,e)\|P(g,e)) \qquad (6)$$

It is important to notice that absolute and relative entropies $H$ and $D_{KL}$ are non-negative metrics. This reconfirms that the highest achievable fitness of all optimal strategies is only achieved with a fitness matrix that is equivalent to Kelly's original diagonal matrix with one surviving type per environmental state. Equation $(6)$ shows the three factors that can reduce this overall optimum. One is the remaining uncertainty of the environment after average updating $H(E|G')$. The second one is a divergence between the hypothetical weighting matrix $P(E|M)$ that intermediates between the real- and a theorized purely diagonal fitness matrix. It turns out that inside the region of bet-hedging the hypothesized weighting matrix $P(E|M)$ is equal to the stochastic matrix $P'(E|G)$, and therefore $D_{KL}(P'(e|g)\|P(e|m))$ is equal to zero inside the region of bet-hedging. The last constrain remains the force of evolutionary updating from our previous equations. We know that the local optima is achieved if this last term is equal to $I(G';E)$. This allows us to create a coherent picture that shows the relation between all of the previous cases.

## Bringing it all together: the big picture.

Table 1 provides a comparative summary of the previously treated cases, plus two additional cases.

Kelly's result is the special case of a noiseless communication channel between the population and the environment. With only one specialized type per environmental state, there is no cross-over noise in the communication channel between the environment and the updating population ($H(E) = H(G')$, see Fig 2b). With a proportional bet-hedging strategy, the last term of our decomposition reduces to the environmental uncertainty $H(E)$ and the divergence $D_{KL}(P(e)||P(g))$, which Kelly proposed to set to zero through proportional bet-hedging[30]. From an information theoretic perspective, this is the maximal value of the mutual information that appears in equation $(4)$: $H(E) \geq I(G';E)$. Shannon[6] termed the maximum of the mutual information $I$ the "channel capacity", the upper bound on the rate at which information can be reliably transmitted over a communications channel. With this finding he had set off a long search for coding schemes to achieve this channel capacity. While it is usually extremely difficult to calculate the channel capacity in practice through numerical algorithms or nonlinear optimization methods, it is in our case revealed by Kelly's criteria with a purely diagonal fitness matrix with $H(E) = I(G';E)$.

With more than one non-fatal type fitness per environmental state, the shape of this mixed fitness landscape introduces constraints that manifest themselves as a noisy communication channel between the environment and the updated population, $H(E|G')$, and in the divergence from this specialized diagonal fitness matrix, $D_{KL}(P'(e|g)||P(e|m))$. Inside the region of bet-hedging it is possible to adjust the bet-hedging strategy according to the distortion of the mixed fitness landscape, setting the latter constraint to zero. Table 1 reveals that there are two ways to think about the remaining terms. One follows Fig. 1 and is represented in equation $(4)$. The other one decomposes the mutual information between the updated population and the environment and cancels out the conditional entropy $H(E|G')$. This

leaves us with the hypothetical diagonal fitness matrix ${}_{hyp}^{d}W$, minus the uncertainty of the environment, $H(E)$, which is equivalent to the population fitness achieved in Kelly's case[33].

Two newly introduced cases in Table 1 refer to cases where all bets are placed on one type. In the first case this is the result of optimization outside the bet-hedging region. In this case[10,11] the best choice is to place all weights on types for which optimization suggests $p(g) \geq 1$, and no bets on the ones with $p(g) \leq 1$. The achievement of full channel capacity is compromised by both the mismatch with the optimal diagonal fitness matrix and the uncertainty about the environment $H(E)$. The second new case in Table 1 shows that the latter goes to zero with a perfect cue that completely describes the dynamic of the unfolding environment. A perfect cue absorbs all environmental uncertainty: $H(E) = H(C)$. In terms of a Venn diagram representation, this would imply a picture similar to the complete overlap shown in Fig. 2a, with the difference that $C$ and $G'$ are switched. From Markovity it follows that in this case the uncertainty of the updated population cannot be smaller than the entropy of the cue, as it is completely absorbed through updating[34]: $H(G') \geq H(E) = H(C)$.

The last case in Table 1 is the case shown we started with. It is a descriptive decomposition of evolutionary change that can be used to describe any kind of evolutionary dynamic. $D_{KL}(P'(g,e)\|P(g,e))$ is an indicator for the force of evolutionary change on a given level of analysis.

**Table 1 | Bringing it all together: the complete decomposition and its special cases.**

| | Long-term population fitness | = | time avg. of corresponding diagonal fitness landscape | environmental uncertainty given avg. updating | divergence between the environment given the mixing matrix and the environment given avg. updating | divergence between the avg. original and avg. updated population |
|---|---|---|---|---|---|---|
| **equ. (6)** | $\log \overline{\overline{W}}$ | = | $E_e\left[\log {}_{hyp}{}^{d}W\right]$ | $-H(E\|G')$ | $-D_{KL}(P'(e\|g)\|\|P(e\|m))$ | $-D_{KL}(P'(g,e)\|\|P(g,e))$ |
| **equ. (3)** | Kelly's case | | $E_e[\log {}^{d}W]$ | 0 | 0 | $-H(E)$<br>$-D_{KL}(P(e)\|\|P(g))$ |
| **equ. (4)** | bet-hedging inside region of bet-hedging | | $E_{p'(g,e)}[\log w(g,e)]$ | | 0 | $-I(G';E)$ |
| | | | $E_e\left[\log {}_{hyp}{}^{d}W\right]$ | ~~$-H(E\|G')$~~ | 0 | $-H(E)+$ ~~$H(E\|G')$~~ |
| **optimal outside bet-hedging region** | | | $E_e\left[\log {}_{hyp}{}^{d}W\right]$ | $-H(E)$ | $-D_{KL}(P'(e\|g)\|\|P(e\|m))$ | 0 |
| **optimal with perfect cue** | | | $E_e\left[\log {}_{hyp}{}^{d}W\right]$ | 0 | $-D_{KL}(P'(e\|g)\|\|P(e\|m))$ | 0 |
| **equ. (2)** | descriptive aggregate | | $E_{p'(g,e)}[\log w(g,e)]$<br>recursively decomposable multilevel term | | | $-D_{KL}(P'(g,e)\|\|P(g,e))$ |

## Empirical proof of concept.

We now calculate some of the above metrics for an empirical case to obtain a feeling for the orders of magnitude of each component and to show how the obtained results can be used normatively to optimize fitness. While the logic of bet-hedging strategies is a common phenomenon of biological evolution and has been identified in the evolution of bacteria[28,35], plants[28,36], insects[37], crustaceans[38], and amphibians[39,40], a solid interpretation of bet-hedging requires the identification of the involved feedback loops that assure a stable share of the evolving population. In biological evolution, the identification of such mechanism for the continuous redistribution of resources is often tricky to find and not generalizable. On the contrary, the social evolution of populations made of conscious entities provides for a large variety of readily identifiable redistribution mechanisms to achieve bet-hedging. For example, redistribution mechanisms among evolving economic sectors include taxes and government

expenditures, insurances, private sector cross-subsidies, trade and capital restrictions, and related kinds of regulation and incentives. For reasons of straightforward interpretability, and data availability, we will therefore work with the easily tractable example of the social evolution of the wood and textile sectors in the manufacturing and handicrafts population of the Swedish economy from 1800 to 1980[41]. The mathematics would of course be analogous if the wood sector would be replaced with an example from biology or ecology, such as the share of woodpeckers in an ecological animal population, or types of trees in a forest population, etc. (evolving through changing shares, as presented in Fig. 3a).

Over the period of 180 years, we obtain an empirical population fitness of $2^{0.0698} \approx 1.04957$, or a compound annual growth rate of some $4.957\ \%$. The application of the descriptive equations $(6)$ and $(2)$ reveal that the average evolutionary force of change through updating, $D_{KL}(P'(g,e)||P(g,e))$, is comparatively small, but incessantly transfers small pieces of valuable information through evolutionary change of the population structure. The environmental uncertainty faced by the evolving population is much larger.

We can now optimize fitness by searching for the channel constellation for which each channel output can be assigned an unequivocally channel input (through mutual information). This is a well-known problem from the engineering of communication channels. In our case, we have seen that this is fulfilled if the following independence condition is fulfilled: $p(g,e) = p'(g) * p(e)$; which implies that the time-average of relative fitness is equal to 1: $E_e\left[\frac{w(g,e)}{\bar{W}(e)}\right] = 1$ (see Methods). Solving for this condition and adjusting the population shares accordingly (Fig.2b) allows us to increase fitness to a compound annual growth of $2^{0.07015} \approx 1.04982$. The increase in fitness represents the fitness value of the information revealed by the identification of the environmental pattern and its corresponding fitness landscapes. It is the information about the environmental patterns that allows us to obtain this increase.

We now suppose that an observer starts to measure fitness as a function of the previous year's weather[42]. Figure 3c reveals that this side information allows to increase fitness to $2^{0.07334} \approx 1.0521$. In the case of

a perfect cue that would deterministically reveal the environmental state, fitness could even be increased to $2^{0.11214} \approx 1.08083$, or a or a compound annual growth rate of $8.083\,\%$, which is almost two thirds higher than the empirically detected population fitness. Comparing the constituents of these last two cases in Fig. 3c reveals that any attempt to get closer to the unattainable noiseless channel can confront a trade-off between environmental uncertainty and constraints due to the mixed fitness landscape. The case of perfect information does not count with any environmental uncertainty, but increasingly faces the constraints of the mixed fitness landscape.

a)


b)

| Fitness matrix | Population type | |
|---|---|---|
| | **Wood & Textile** | **Rest of economy** |
| **W&T friendly environment** 46.7 % | **1.0848** ${}^{d}_{hyp}W = 2.227$ | 1.0188 ${}^{d}_{hyp}W = 0$ |
| **W&T hostile environment** 53.3 % | 1.0185 ${}^{d}_{hyp}W = 0$ | **1.0774** ${}^{d}_{hyp}W = 1.986$ |

c)

| **equ. (6)** | **Log of population fitness** | **=** | **noiseless channel** | **environ. uncertainty** | **fitness landscape constraint** | **evolutionary force of change** |
|---|---|---|---|---|---|---|
| **equ. (6) descriptive** | 0.06980 | = | 1.06694 | - 0.99663 | - 0.00009 | - 0.00042 |
| **equ. (2) descriptive** | 0.06980 | = | 0.07022 | | | - 0.00042 |
| **equ. (4) optimal in bet-hedging region** | 0.07015 | = | 0.07069 | | | - 0.00054 |
| | | | 1.06694 | - 0.99625 | 0 | - 0.00054 |
| Given the cue of the previous year's temperature (outside of bet-hedging region): | | | | | | |
| **equ. (5) optimal outside bet-hedging region** | $\log \bar{\bar{W}}_{\mid c}$ | = | $E_e[\log {}^{d}_{hyp}W]$ | $-H(E\mid C)$ | $-E_c[D_{KL}(P'(e\mid g)\Vert P(e\mid m))]$ | 0 |
| | 0.07334 | = | 1.06694 | - 0.99066 | - 0.00295 | 0 |
| **optimal with perfect cue** | 0.11214 | = | 1.06694 | 0 | - 0.95480 | 0 |

**Figure 3 | Evolution of the Swedish economy from 1800 - 1980.**

**Figure 3 | Evolution of the Swedish economy from 1800 - 1980. (a)** Visualization of the corresponding evolution of population types from 1800 to 1980. **(b)** We distinguish between two environmental states: a wood & textile friendly environment, which occurs with $\mathrm{p(e)} = 0.467$ during the 180 years of our sample, and an environment in which this type grows below average fitness. The resulting fitness matrix is shown. We now optimize fitness by setting the time-average of relative fitness for one type to 1, for example, for the wood & textile type: $E_e\left[\frac{w(g,e)}{\bar{W}(e)}\right] = 0.467 * \frac{1.0848}{\bar{W}(e=W\&T_{friendly})} + 0.533 * \frac{1.0185}{\bar{W}(e=W\&T_{hostile})} = 1$. A numerical (and/or visual) solution under the given constraints provides $\bar{W}(e = W\&T_{friendly}) = 1.0393$ and $\bar{W}(e = W\&T_{hostile}) = 1.0592$, which is fulfilled for a stable share of $p(g = W\&T) \approx 31\ \%$ and $p(g = Rest) \approx 69\ \%$. For example: $0.31 * 1.0848 + 0.69 * 1.0188 = 1.0393$. Maintaining this distribution of types assures optimal population fitness. We can now also calculate the hypothetical noiseless channel equivalent, with ${}_{hyp}^{d}W(e = W\&T) = 2.227$ and ${}_{hyp}^{d}W(e = Rest) = 1.986$ (see Fig. 3b), which leads to a channel capacity of $E_e[\log {}^{d}W] = 1.06694$. **(c)** Shows the quantities represented by the decomposition for various cases. For the case with side information (equation $(5)$), the observers distinguishes between an annual temperature above and below the Northern European average of $9.27\ ℃$ . Conditioned on higher or lower realizations of this random variable, two different fitness landscapes are recorded. This results in two different bet-hedging strategies. Unfortunately, for both cases, the optimal strategy falls outside the region of bet-hedging, which creates a constraint. The fitness is optimized when all resources are bet on the type of wood and textile in the years in which temperature is above average, and nothing on this type in other years.

## Discussion

A typical dictionary definition of evolutionary fitness is twofold: "fitness (biology): (a) the degree of adaptation of an organism to its environment…; (b) the ability of an organism to produce viable offspring…"[43]. The decompositions presented here show that both intuitive definitions can be linked through an information theoretic logic. Fitness can be expressed in terms of the informational fit between environmental patterns in time, and population structures in space. This results in a communication process, whereas the role of the sender in a particular exchange defines the role of the environment (roles can be switched).

As such, the presented decomposition are generalizations of an emerging body of recent literature that has suggested that information theory can be useful for describing evolutionary fitness[3 - 17]. Information theory is arguably the scientific theory with the most widespread practical impact on our daily lives. Shannon's ideas of entropy (compression) and mutual information (channel capacity) led to all kinds of digital communication technology, which have become indispensable for humanity within only a couple of decades. This is an impressive achievement for a scientific theory. The application of these powerful ideas to evolutionary dynamics leads to an open research agenda. On the one hand, our explorations of fitness merely involved the question of simple description and optimization, leaving open a vast array of interesting questions yet to be explored. On the other hand, the information theoretic concepts that emerge from our reformulations are merely the most fundamental metrics, including absolute and relative entropies and mutual information. Information theory is a well-established scientific field that goes way beyond these basic metrics. For example, the current standard textbook on information theory[1] treats these fundamental metrics in the first substantive Chapter (p. 13-40), followed by 15 additional Chapters with additional 650 pages. Both of these facts combined gives a basic notion of the magnitude of the arising research agenda on the analysis of evolutionary dynamics from an information theoretic perspective.

One of the aspects of this agenda relates to questions regarding the space-time nature of the informational definition of fitness, and, closely linked, to the omnipresent question of the adequate level of selection. The benchmark of Kelly's diagonal fitness matrix suggests a logic of space-time requisite variety, whereas specialized types correspond to different environmental states. Additionally, the nature of equation $(2)$ reveals a recursive multilevel logic over multiple levels of a taxonomy. This leads to a new look on the question of the adequate levels of selection and how space-time fine-graining defines a level. Another unanswered question refers to the fact that the derived results underline the importance of bet-hedging as a benchmark in evolutionary fitness. Bet-hedging requires some kind of negative feedback

mechanisms that constantly redistributes resources in order to maintain the population distribution. In essence, bet-hedging 'stops' natural selection and leads to stable population shares[44]. In the dynamics of blind neo-Darwinian selection, these redistributing feedback mechanisms have been discovered through random mutations, which obtained an evolutionary advantage at the higher level (based on the superiority of bet-hedging). In other words, given stable shares of types on the lower level, natural selection can thrive on the higher level and promote these kinds of mechanisms. Bet-hedging is common in a wide variety of plant and animal species[35 - 40]. Redistribution mechanisms are more straightforward and common in rather Lamarckian notion of social evolution with conscious foresight[28 - 28], including portfolio management, tax and subsidies, counter-cyclical economic policy, and a wide variety of long-term planning tools. The logic presented here formalizes the intuitive fact that the obtainable growth rate hinges on the amount of information available about the environment. Less uncertainty, more growth.

Last but not least, it is important to note that most of currently existing information theory is based on the notions of ergodicity and stationarity. Applied to our case it is easy to see that the identification of an environmental pattern requires the existence of a stable pattern. It might turn out that these limitations are more severe in the evolution of biological and social systems in comparison to engineering problems, the main field of traditional applications of information theory. For example, we know that fitness landscapes are density dependent and change with the evolving population. Dealing with non-ergodic and non-stationary dynamics might require extensions to the existing tool box of information theory.

# Methods

## Basic decomposition.

Two main steps are involved in the reformulations that result in equations $(1)$ and $(2)$. First, the expected value is taken on $\overline{\overline{W}}$ (or its log), which is justified by the fact that the expected value of a constant is the same constant (e.g. $\log \overline{\overline{W}} = E\left[\log \overline{\overline{W}}\right]$). Second, we employ a revers form of the so-

called replicator equation to decompose average population fitness per environment into lower level type fitness:

$$\log \bar{\bar{W}} = E_{p'(g|e)}\left[\log \bar{\bar{W}}\right] = \sum p'(g|e)\left[\log\{\bar{W}(e)\}^{p(e)}\right] = \sum p'(g|e)p(e)\left[\log\left\{w(g,e)\frac{p(g|e)}{p'(g|e)}\right\}\right]$$

$$= \sum_{g,e} p'(g,e)[\log w(g,e)] - \sum_{g,e} p'(g,e)\log\frac{p'(g,e)}{p(g,e)} = \text{equ. (2)}$$

## Complete decomposition.

The reformulation of equation $(6)$ is obtained by replacing the true fitness values $w(g,e)$ in equation $(2)$ with the weighted hypothetical diagonal fitness values, and then expanding with term with $p'(g,e)$.

$$\log \bar{\bar{W}} = E_{p'(g,e)}\left[\log\left(\left\{{}^{d}_{hyp}W(e) * p(e|m)\right\} * \frac{p'(g,e)}{p'(g,e)}\right)\right] - D_{KL}(P'(g,e)\|P(g,e)) = \text{equ. (6)}$$

## Optimality of $D_{KL} = I$.

The condition that $D_{KL}(P'(g,e)\|P(g,e)) = \sum_{g,e} p'(g,e)\log\frac{p'(g,e)}{p(g,e)}$

$= \sum_{g,e} p'(g,e)\log\frac{p'(g,e)}{p'(g)*p(e)} = I(G';E)$ is fulfilled for $p(g,e) = p'(g) * p(e)$, since the mutual information is defined as the relative entropy between the joint distribution and the corresponding independent distribution. With the help of the replicator equation, this can be rewritten as $p(g|e) = p'(g) = \sum_e p(e) * p'(g|e) = \sum_e p(e) * p(g|e) * \frac{w(g,e)}{\bar{W}(e)}$, whereas $p(g|e)$ cancels out and we obtain the condition that the time-average of relative fitness is equal to 1:

$$\sum_e p(e)\frac{w(g,e)}{\bar{W}(e)} = 1 \qquad (7)$$

The two-way proof that this implies optimal growth and vice versa is easily shown when working with the hypothetical fitness matrix of the channel capacity from the section The complete decomposition with constraints. First we show that optimal fitness implies the existence of mutual information in our decomposition: $optimality \Rightarrow D_{KL} = I$. We express both the numerator and denominator of equation $(7)$ with their equivalent expressions from the full channel capacity fitness matrix:

$$\sum_e p(e)\frac{w(g,e)}{\bar{W}(e)} = \sum_e p(e)\frac{p(e|m_i) * {}^{d}_{hyp}W(e)}{p(e) * {}^{d}_{hyp}W(e)} = \sum_e p(e) * p(e|m_i) = 1$$

Whereas the last step follows from stochasticity of the weighting matrix (in this case of optimal bet-hedging in mixed fitness landscapes $p(e|m_i) = p(e|g_i)$).

Two additional assumption are required for: $D_{KL} = I \Rightarrow optimality$. The first one are stable shares of types (either through bet-hedging or through any other kind of stable equilibrium in the population shares), and the second one is that population types and environmental shares are linearly independent.

We reformulate the stochastic matrix $1 = \sum_e p(e|m_i) = \sum_e \frac{w(g_i,e)}{{}^{d}_{hyp}W(e)}$. At this point we include the restriction of the region of bet-hedging, which is $0 \leq p'(g|e) \leq 1$, and replace $w(g_i, e)$ with the reverse form of the replicator equation. $\sum_e \frac{p'(g_i|e)}{p(g_i|e)}\bar{W}(e)\frac{1}{{}^{d}_{hyp}W(e)} = \sum_e \frac{p'(g_i,e)}{p(g_i|e)}\ \frac{\bar{W}(e)}{p(e)\ {}^{d}_{hyp}W(e)} = 1$. We assume that $p(g_i|e)$ is fixed for a specific type $i$ (e.g. through bet-hedging) (which allows us to bring it to the right hand side) and that $D_{KL} = I$, which implies $p(g_i, e) = p(g_i|e) * p(e) = p'(g_i) * p(e)$, or $p(g_i|e) = p'(g_i)$. This leaves us with $\sum_e p'(g_i, e)\frac{\bar{W}(e)}{p(e)\ {}^{d}_{hyp}W(e)} = p'(g_i) = \sum_e p'(g_i, e)$. We can rewrite this in matrix form for all types $i$ over all types and environmental states:

$$\mathbf{P} * \vec{w} = \overrightarrow{p'}$$

$$\begin{bmatrix} p'(g=1,e=1) & p'(g=2,e=1) & \dots \\ p'(g=1,e=2) & \dots & \dots \\ \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} \bar{W}(e=1)/\left(p(e=1)\ {}^{d}_{hyp}W(e=1)\right) \\ \bar{W}(e=2)/\left(p(e=2)\ {}^{d}_{hyp}W(e=2)\right) \\ \dots \end{bmatrix} = \begin{bmatrix} \sum_e p'(g_i,e=1) \\ \sum_e p'(g_i,e=1) \\ \dots \end{bmatrix}$$

If the rank of the coefficient matrix $\mathbf{P}$ is equal to the rank of the respective augmented matrix $\mathbf{P}^{\#}$, the system is consistent and must have at least one solution (Rouché–Capelli theorem [45]). This is the case here, since the last column of the augmented matrix, $\overrightarrow{p'}$, can easily be set to 0s through column operations of $\mathbf{P}$ (which do not affect the rank; i.e. subtracting each column once). So whatever the rank of $\mathbf{P}$ will be the rank of $\mathbf{P}^{\#}$**.** We furthermore know that the solution is unique if the rank is equal to the number of variables. Otherwise we have infinitely many solutions [45]. The trivial case for the condition of

a unique solution is Kelly's diagonal fitness matrix, with non-zero values only in the diagonal, which is already in reduced echelon form. In the case that either different types or different environments are linearly dependent we obtain infinitely many solutions (dependence refers to the concept from linear algebra here, not to random variables). If these redundant states and types are merged, the number or variables is equal to the rank. To identify the unique solution, we work with a method that works for many such problems: guess and verify. It is straightforward to show that the unique solution to the system is $\frac{\bar{W}(e)}{p(e)\ {}_{hyp}^{d}W(e)}$ for all environments, which shows optimality.

## References

[1] Cover, T. M. & Thomas, J. A. *Elements of Information Theory*. (Wiley-Interscience, 2006); p. 183.

[2] Between 1924 and 1934, J.B.S. Haldane published a series of ten papers (mainly in the *Proceedings of the Cambridge Philosophical Society*) called "A Mathematical Theory of Natural and Artificial Selection", which outlines the first mathematical models for many cases of evolution due to selection; see e.g. Haldane, J. B. S. A mathematical theory of natural and artificial selection—I. *Transactions of the Cambridge Philosophical Society* 23, 19–24 (1924).

[3] Frank, S. A. *Foundations of Social Evolution*. (Princeton University Press, 1998). at <http://stevefrank.org/foundations/foundations.html>

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

[5] Nelson, R. R. & Winter, S. G. *An Evolutionary Theory of Economic Change*. (Belknap Press of Harvard University Press, 1985).

[6] Shannon, C. A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 379–423, 623–656 (1948).

[7] MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge, UK ; New York: Cambridge University Press.

[8] Frank, S. A. Natural selection maximizes Fisher information. *Journal of Evolutionary Biology* 22, 231–244 (2009).

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

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

[11] Donaldson-Matasci, M. C., Bergstrom, C. T. & Lachmann, M. The fitness value of information. *Oikos* 119, 219–230 (2010).

[12] Crutchfield, J. P. & Schuster, P. *Evolutionary Dynamics: Exploring the Interplay of Selection, Accident, Neutrality, and Function*. (Oxford University Press, 2003).

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

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

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

[16] Krakauer, D. C. Darwinian demons, evolutionary complexity, and information maximization. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 21, 037110–037110–12 (2011).

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

[18] Neumann, J. V. & Morgenstern, O. *Theory Of Games and Economic Behavior*. (Princeton University Press., 1944). at <http://archive.org/details/theoryofgamesand030098mbp>

[19] Fisher, R. A. *The Genetical Theory of Natural Selection*. (Clarendon Press, Oxford, 1930).

[20] Crow, J. F. & Kimura, M. *An introduction to population genetics theory*. (Blackburn Press, 1970).

[21] Kullback, S. & Leibler, R. A. On Information and Sufficiency. *The Annals of Mathematical Statistics* 22, 79–86 (1951).

[22] Sanov, I. N. On the probability of large deviations of random variables. *Mat. Sbornik* 42, 11–44 (1957); transl. (1961).

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

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

[25] Latané, H. A. Criteria for Choice Among Risky Ventures. *Journal of Political Economy* 67, 144–155 (1959).

[26] Blume, L. E. & Easley, D. Optimality and Natural Selection in Markets. *Journal of Economic Theory* 107, 95–135 (2002).

[27] Hens, T. & Schenk-Hoppe, K. R. Evolutionary finance: introduction to the special issue. *Journal of Mathematical Economics* 41, 1–5 (2005).

[28] Seger, J. & Brockmann, J. in *Oxford Surveys in Evolutionary Biology* 4 (eds. Harvey, P. H. & Partridge, L.) 182–211 (Oxford University Press, 1987). at <http://biologylabs.utah.edu/seger/Seger_Brockmann_87.pdf>

[29] Rainey, P. B. et al. The evolutionary emergence of stochastic phenotype switching in bacteria. *Microb. Cell Fact. 10th Symposium on Lactic Acid Bacterium*, S14 (2011).

[30] For a good overview of Kelly's result in information theoretic terms see Chapter 6 in reference 1.

[31] James, R. G., Ellison, C. J. & Crutchfield, J. P. Anatomy of a bit: Information in a time series observation. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 21, 037109 (2011).

[32] Yeung, R. W. A new outlook on Shannon's information measures. *IEEE Transactions on Information Theory* 37, 466–474 (1991).

[33] This natural result of our approach was the starting point of the analysis in references 10 and 11.

[34] This follows from the definition of mutual information and the data processing inequality (see references 1 and 7: $H(G') \geq I(G';E) \geq I(G';C) = H(E) = H(C)$.

[35] Maamar, H., Raj, A. & Dubnau, D. Noise in Gene Expression Determines Cell Fate in Bacillus subtilis. *Science* 317, 526–529 (2007).

[36] Cohen, D. Optimizing reproduction in a randomly varying environment. *J. Theor. Biol*. 12, 119–129 (1966).

[37] Danforth, B. N. Emergence dynamics and bet hedging in a desert bee, Perdita portalis. *Proceedings of the Royal Society of London B: Biological Sciences* 266, 1985–1994 (1999).

[38] Philippi, T., Simovich, M., Bauder, E. & Moorad, J. Habitat Ephemerality and Hatching Fractions of a Diapausing Anostracan (crustacea: Branchiopoda). *Israel Journal of Zoology* 47, 387–396 (2001).

[39] Crump, M. L. Variation in Propagule Size as a Function of Environmental Uncertainty for Tree Frogs. *American Naturalist* 117, 724–737 (1981).

[40] Morey, S. R. & Reznick, D. N. The relationship between habitat permanence and larval development in California spadefoot toads: field and laboratory comparisons of developmental plasticity. *Oikos* 104, 172–190 (2004).

[41] Data from Edvinsson, R. *Growth, Accumulation, Crisis : With New Macroeconomic Data for Sweden 1800-2000*. (Acta Universitatis Stockholmiensis, 2005). at <http://su.diva-portal.org/smash/record.jsf?pid=diva2:193148>

[42] Met Office. *Met Office Hadley Centre Central England Temperature Data*: Monthly_HadCET_mean, 1659 to date. (Met Office Hadley Centre for Climate Change, 2015). at <http://hadobs.metoffice.com/hadcet/data/download.html>

[43] Farlex, I. N. C*. The free dictionary*. (2015). Also: "Fitness (Symbol W): A measure of the relative breeding success of an individual or genotype in a given population at a given time. Individuals that contribute the most offspring to the next generation are the fittest. Fitness therefore reflects how well an organism is adapted to its environment, which determines its survival". Martin, E. & Hine, R. *A Dictionary of Biology*. (Oxford University Press, 2008). at <http://www.oxfordreference.com/view/10.1093/acref/9780199204625.001.0001/acref-9780199204625>

[44] Hilbert, M (forthcoming). A formal model of how structural change hinges upon knowledge and information. *Structural Change and Economic Dynamics*.

[45] See any textbook on linear algebra, such as Goode, S. W. *Differential Equations and Linear Algebra*. (Prentice Hall, 1999); or Meyer, C. D. *Matrix Analysis and Applied Linear Algebra*. (SIAM: Society for Industrial and Applied Mathematics, 2001).

## Acknowledgements

I am indebted with Steve Frank for continuously encouraging me to look deeper, and with Jim Crutchfield and his team for continuously exposing me to the depth and beauty of information theory, as well as to Matina Donaldson-Matasci, Ryan James, Gerhard Kramer, Poong Oh, Peter Monge, and the faculty of the Santa Fe Institute for direct or indirect discussions resulting in this paper.