

Formal definitions of information and knowledge and their role in growth through structural change

Martin Hilbert
University of California, Davis; hilbert@ucdavis.edu

Author's version, published as: Hilbert, M. (2016). Formal definitions of information and knowledge and their role in growth through structural change. *Structural Change and Economic Dynamics*. <http://doi.org/10.1016/j.strueco.2016.03.004>

Abstract

The article provides a way to quantify the role of information and knowledge in growth through structural adjustments. The more is known about environmental patterns, the more growth can be obtained by redistributing resources accordingly among the evolving sectors (e.g. bet-hedging). Formal equations show that the amount of information about the environmental pattern is directly linked to the growth potential. This can be quantified by treating both information and knowledge formally through metrics like Shannon's mutual information and algorithmic Kolmogorov complexity from information theory and computer science. These mathematical metrics emerge naturally from our evolutionary equations. As such, information becomes a quantifiable ingredient of growth. The policy mechanism to convert information and knowledge into growth is structural adjustment. The presented approach is applied to the empirical case of U.S. export to showcase how information converts into growth potential.

Keywords: evolutionary growth, information, international trade, bet hedging, fitness decomposition, knowledge.

JEL codes: B25; B52; C02; D80; D81; D92; F14; G11; O43.

Highlights:

- Computer science definitions of information and knowledge are applied to economics
- Shannon entropy and mutual information emerge from our evolutionary equations
- Information and knowledge become quantifiable ingredients of growth
- A direct relationship is shown between information and growth potential
- Changing economic structures according to this information optimizes growth

Graphical Abstract:



The Nobel-laureate and co-founder of the Santa Fe Institute, Murray Gell-Mann, came to the conclusion that although complex systems “differ widely in their ... attributes, they resemble one another in the way they handle information. That common feature is perhaps the best starting point for exploring how they operate” (1995, p. 21). The article uses formal definitions of information and knowledge from information theory and computer science, links them to related definitions from evolutionary economics, and showcases the how information can inform structural change in the economy. Information theoretic metrics of information naturally emerge from evolutionary decompositions of growth and can directly be linked to the growth potential due to the redistribution of resources (i.e. structural change of the economic population).

An intuitive micro-economic example will set the stage and introduce the presented argument. Imagine a bakery that produces salty and sweet goods. If nothing else is known about the future environment, economic evolution would suggest to let market selection winnow out the more profitable among these two business options (Nelson and Winter, 1985). However, if we have information about future dynamics, it might be profitable to intervene into market selection. Recent big data analysis of statistical patterns has revealed that the demand for sweet goods grows with rain and the demand for salty goods with sunshine. This information about the environment allows to adjust the structure of its product offerings to the identified environmental pattern. The more it knows about the relation, the more growth potential. Being aware of this relationship and the environmental pattern with regard to rain and sunshine, productivity increments of up to 20 % have been reported for individual bakeries (Christensen, 2012). The potential to grow depends on what is known about the uncertain future environment.

The article shows how information and the resulting growth potential relate through a logic of betting. This includes two steps. “On the one hand there is a need to observe and to know, on the other hand there is a need to modify the external environment” (Saviotti, 2004; p. 101). Or to use the well-known concepts from Lundvall and Johnson (1994): for one we need to ‘know what’ (e.g. the relation between baked goods and the weather), and additionally we need to ‘know how’ (e.g. how to redistribute proportions to optimize growth). We will use a simple evolutionary decomposition of growth to formalize the optimal relationship between growth and the distribution of the evolving population (be it proportions of baked goods or shares of economic sectors). We will forgo the details of how this redistribution can be implemented in practice (as there is a large variety of options available), but focus on the fact that there is a formal and quantifiable relation between the amount of information and the potential to grow. The more is known about the future, the more efficient can be the restructuring, the higher the obtainable growth rate. Better information and the respective redistribution of resources hinge upon each other. This converts formal definitions of information and knowledge into a quantifiable input for economic growth. Longstanding information theoretic metrics of information emerge as an integral part of our evolutionary equations, including Shannon’s (1948) entropy and mutual information, Kullback-Leibler’s relative entropy (1951), Massey’s directed (causal) information (1990), and Kolmogorov’s algorithmic information / knowledge (i.e. Kolmogorov (1968) complexity). These metrics are traditionally used in engineering, physics and computer science, but turn out to be very useful when conceptualizing economic dynamics. The policy tool to convert information and knowledge into growth is structural change through the redistribution of resources among the constituent sectors of an evolving economic population.

A formalization of the relation between information and knowledge and growth becomes increasingly important in a big data world in which a myriad of previously unknown environmental patterns are

identified thanks to the unprecedented amount of available information about all kinds of dynamics and structures (e.g. Mayer-Schönberger and Cukier, 2013; Hilbert, 2016). It is claimed that the “data-driven economy” (European Commission, 2014) thrives on plain “data as a new source of growth” (OECD, 2013). The article shows how formal notions of information and knowledge extracted from data relate to growth through structural change, which formalizes the idea of “harnessing big data to unleash the next wave of growth” (Manyika et al., 2011).

The article consists of four main parts. It first introduces the formal metrics used to quantify information and knowledge, such as “Kolmogorov complexity” and “Shannon information”. References are made to related conceptualizations from the literature of evolutionary economics and business knowledge management to put these mathematical notions into the existing social context. The next section uses a complete decomposition of growth and introduces the logic of bet-hedging. Bet-hedging can be shown to optimize growth given existing information / knowledge about environmental patterns. This follows the basic result of Kelly (1956) and more recent contribution in the field of portfolio theory and biological bet-hedging. The third section provides empirical evidence of the derived results. It quantifies the role of information for potential structural adjustments in the U.S. export economy. The final sections discuss the underlying assumptions of the presented logic and possible extensions.

1. How to formalize “knowledge” and “information”?

We start by defining our notions of information and knowledge. There is a desperate need for “sharpening the distinctions between information and knowledge... [last but not least, due to] the continued acceleration of innovations in information and communication technologies” (Cohendet and Steinmueller, 2000; p. 195). We will turn the tables of this momentum and use the proper definitions of information and knowledge employed by information and communication technologies to distinguish between them. In agreement with the well-known data-information-knowledge framework from the business knowledge management literature (Zeleny, 1986; Ackoff, 1989), also in information theory “information is defined in terms of data, knowledge in terms of information” (Rowley, 2007; p. 163). Also engineers and computer scientists see “information and knowledge are two very different, although related concepts” (Saviotti, 2004; p. 105). It turns out that formal definitions of knowledge (in form of “Kolmogorov complexity”) and information (in the form of “Shannon entropy”) are two sides of the same coin. Just like two sides of a coin, they are not identical, but they are two complementary ways of showing the same average quantity.

1.1 Data, information, and knowledge in information theory

In information theory, *data* is usually defined as symbols without any meaning or value. Data is defined as *any distinction, any perceivable difference*. This variation can be visual, auditory, tactile, olfactory, gustatory, imagery, or dynamic in time, etc. Digital data are represented as different kinds of differences, the most fundamental of them being binary, such as current or no current (in an electronic circuit), light or no light (in a fiber-optic cable), certain wave or not (in cellular frequency spectrum), etc.

Conceptually the placeholder 0 and 1 is often taken for this most fundamental binary difference, but this notation is simply social practice and can be replaced with any exclusive and exhaustive difference.

Information is a “difference which makes a difference” (Bateson, 2000; p. 272). The difference that it makes is that it *reduces uncertainty*. If some kind of symbol shows a difference (for example either 1 or 0), but this does not represent any surprise, it does not reduce uncertainty, and therefore is no information. It is simply some kind of redundant data. Information can be extracted from data by ‘compression’, which essentially takes out all redundant data and leaves only those differences that effectively reduce uncertainty. The remaining quantity is called the “entropy of an information source” (Shannon, 1948). This is Shannon’s famous “source coding theorem” (Cover and Thomas, 2006). So information is defined as the opposite of uncertainty, which makes intuitively sense. If you have uncertainty, you do not have information, and the communication of information is the process of uncertainty reduction. Since uncertainty can be quantified with the mathematical tools of probability theory, Shannon’s information can be quantified as well in probabilistic terms. Shannon defined *one bit* of information to be ‘that that *reduces uncertainty by half*’. Bits (in Shannon’s sense) are the fundamental information units. They are defined in terms of uncertainty, not in terms of how many data symbols exist. A binary digit is only an information bit if it reduces uncertainty by half. This can be related to the somewhat vague but very common definition that information is “data that have been organized... [or] processed... [or] interpreted... [or] given meaning” (Rowley, 2007; p. 171), but is much more precise in its effect and quantitative representation.

In the general epistemology, knowledge is often defined as “information that have been organized” (Rowley, 2007; p. 172) or “as actionable information” (p. 175). While “information is to be interpreted as factual... knowledge, on the other hand, establishes generalizations and correlations between variables. Knowledge is, therefore, a correlational structure” (Saviotti, 1998; p. 845). This implies that knowledge consists of an interlinked chain of information. Nelson and Winter (1985) famously argue that it is in its routines that a firm’s organizational knowledge is stored (also Becker, 2003); psychologists argue that behavioral patterns contain knowledge about the world that allow us to make relevant decisions (Tversky and Kahneman, 1974), and anthropologists claim that cultural norms represent accumulated knowledge (Boyd and Richerson, 2005).

Adopting a strict computer science definition of knowledge, *knowledge* implies a step-by-step recipe of doing something, a deterministic algorithm (e.g. Brookshear, 2009; p. 205). If there is knowledge, there is no doubt, there is a deterministic process or procedure that defines how things go from one to the next. This deterministic process might be unknown (i.e. ‘tacit’, see below), but it is still deterministic. Theoretically, the right measure to quantify the amount of knowledge contained in a deterministically unfolding process has been established by Andrey Kolmogorov (1941; 1968), and independently developed by Solomonoff (1964) and Chaitin (1966). It is known as “Kolmogorov complexity” (Li and Vitanyi, 2008), or any combination of the foregoing three names (Crutchfield, 2012). It defines the amount of ‘knowledge’ in a deterministic algorithm to be *equal to the minimum number of symbols needed to efficiently describe an object to a specific level of detail*. In this sense, algorithms relate symbols through a step-by-step structure, often in time, such as through the algorithmic ‘if-then’ logic. This is in line with Saviotti’s intuitive definition of “knowledge as co-relational structure” (2004; p. 102). The difference to information is that the symbols are not probabilistic (such as information bits), but deterministic. They give instructions, no surprise involved.

1.2 An 'amazing' and 'beautiful' fact

It turns out that Shannon's probabilistic- and Kolmogorov's deterministic approaches are equivalent asymptotically. They are two sides of the same coin. The formal mathematical proof is rather subtle (see Kolmogorov, 1968; Zvonkin and Levin, 1970; Leung-Yan-Cheong and Cover, 1978; Zurek, 1989), but is intuitive and elicits even the most hardened theorists exclamations like "amazing" (Cover and Thomas, 2006; p. 463) and "beautiful" (Li and Vitanyi, 1997; p. 187).

The proof shows that on average and over long descriptions the amount of uncertainty reduction that results from Shannon probabilistic setup is equivalent to the amount of information required by Kolmogorov's approach to describe the object.). Shannon's probabilistic approach is like answering binary questions about the location of a destination on a map ('Is it North or South?'; 'Is it East or West?', 'Is it East or West at the crossing?'). This is in agreement with the epistemological idea that "information is contained in answers to questions" (Bernstein, 2011; also Ackoff, 1989). More precisely, one bit resolves a uniformly distributed binary uncertainty. Kolmogorov's algorithmic approach is like describing the route to a destination on a map (e.g. 'Go North, then West, if at crossing, then turn East, etc.'). The theorem says that both ways of defining the final destination requires on average the same number of symbols. This is essentially because require to create a contrast against the number of alternative possible choices (Cover and Thomas, 2006). The city of Davis in Northern California can be identified by isolating it from within a probability space that selects it from all other cities (most fundamentally by a series of binary choices that reduce uncertainty by half), or it can be identified by giving deterministic instructions of how to bypass all other cities to end up in Davis (most efficiently by giving maximal explanatory power with every instructional symbol). This is the basic gist behind the underlying mathematical proof (for the formal proof see Cover and Thomas, 2006; or Li and Vitanyi, 1997).

Being two sides of the same coin, both quantities also complement each other. For example, when studying complex systems, part of the systems dynamic can be described in terms of deterministic structure and the other part terms of probabilistic uncertainty. This duality is succinctly expressed in familiar concepts of modern complexity science, such as 'algorithmic randomness' (Zurek, 1989), 'statistical mechanics' and 'deterministic chaos' (Crutchfield, 2012). These terms imply that some part of system's dynamic is algorithmic/mechanical/deterministic, while at the same partially being random/statistical/chaotic. One part refers to the part that is known about the system (deterministic, without doubt), the other one to the part that is unknown (probabilistic, in the worst case maximum entropy/uncertainty) (Shalizi and Crutchfield, 2001; Crutchfield and Feldman, 2003).

The same logic is at the heart of the argument that finally exorcised Maxwell's (1872) notorious demon by the hands of Szilard (1929), Bennett (1982) and Zurek (1989). Zurek (1989) showed that the demon can either 'know' the position of a particle to extract energy (through an algorithm), or it can make an uncertainty reducing observation (through informational bits). Both have the same net effect.

Also digital technologies make complementary use of both. For example, in order for the sending and receiving agent to 'know' that an incoming symbol has a 50%-50% chance (or a 20%-80% chance, etc.), it requires a 'lookup table' with the corresponding probabilities. At the very least, the definition of uncertainty always requires a normalized probability space (normalizing probabilities between 0 and 1). In other words, every probability is conditioned on its underlying probability space (defining the number

of possible events and their probabilities). In modern digital applications these encoding and decoding 'lookup tables' are an integral part of technological standards like ZIP, MPEG, JPEG, MP3, CDMA, or UMTS, ATSC, etc. Only through this 'lookup table' can the sender distinguish plain data from uncertainty-reducing information and choose to only send information (through compression), while neglecting redundant data. The existence of the 'lookup table' requires code in form of an algorithm. This algorithm might create a dynamic lookup table (such as with Lempel-Ziv compression), but it is nonetheless an algorithm that can be quantified in terms of Kolmogorov complexity (for a nice theoretical treatment of this complementarity see Caves, 1990). For example, first you have to know that there are 4 events that are equally likely (this is the given 'lookup table'). Then two consecutive bits can allow to identify the chosen one of the 4 choices.

The complementary nature of both is also an integral part of most existing social science definitions of information and knowledge. Saviotti (1998) explains that "particular pieces of information can be understood only in the context of a given type of knowledge" (p. 845); Jensen et al. (2007) point out that "in order to understand messages about the world you need to have some prior knowledge about it (p. 681); and Cowan et al. (2000) explain that "it is the cognitive context afforded by the receiver that imparts meaning(s) to the information-message... The term 'knowledge' is simply the label affixed to the state of the agent's entire cognitive context" (p.216). While these notions invoke sophisticated analogies, in the simplest case, this 'cognitive context' is a simple 'lookup table' (a dictionary or codebook) that allows to distinguish between data and uncertainty-reducing information and to de-codify the incoming symbols. "Thus, initial codification activity involves creating the specialized dictionary" (Cowan et al., 2000; p.225).

Their complementarity has also economic relevance. Algorithms allow to infer information that is part of the algorithmic structure without the need for further measurement (observation). This is advantageous for prediction of the future (where empirical observations are impossible) or "if calculation costs are lower than measurement costs" (Saviotti, 2004; p. 104). In this sense, an economic agent can choose to obtain the information 'bit by bit' through observation (e.g. when it is enough to react to the clouds in the sky) or invest in creating an algorithm that allow for predicting the clouds in the sky. While on average the same amount of information and knowledge are required, one can option be more valuable than the other for particular cases.

1.3 Additional characteristics

There are certain additional characteristics of information and knowledge that usually attract the attention of social scientists and economists. They refer to aspects like value and meaning, tacitness, and the recombinant nature of knowledge. This sections puts our definitions into this context before moving on to the application of our definition of economic growth. As this context is complementary, the rushed reader can safely skip this section.

Meaning and value: Note that none of our information theoretic definitions involve meaning or value per se. Metrics like entropy are always about something, as they quantify a random variable that presents some distribution. For example, let E be a random variable with i realizations, distributed according to $P(E)$; then $H(E)$ refers to the entropy that quantifies the uniformity of the distribution through $H(E) = -\sum_i e_i \log(e_i)$. Analogously, and algorithm always does something. For example, it prioritizes content of a social network or moves a robotic arm. Additional information theoretic metrics also allow to see how much one random variable or one algorithm has in common with another one. These metrics are called

‘mutual information’ (Shannon, 1948; Cover and Thomas, 2006) and ‘information distance’, respectively (Bennett et al., 1998). For example, mutual information allows to quantify how much information a cue C contains about the environment $I(E; C)$ (this metric will crop up in our evolutionary equations below). However, it does not provide deeper meaning or value. Using Shannon’s own words, meaning would imply that the involved symbols (be it plain data, uncertainty reducing information, and sequential knowledge) “refer to or are correlated according to some system with certain physical or conceptual entities” (Shannon, 1948; p. 379). The resolution of a 50-50 chance could resolve an entertaining brain teaser or win you millions of dollars. It would in both cases represent one bit of information. If the bit would be perfectly correlated with another reflective bit, they would have a mutual information of one bit. But this would not give meaningful value to the bit. Likewise, two algorithms with the same number of symbols could enable you to solve humanity’s hunger problem or simply allow you to get dinner from the fridge and heat it up. Both require knowledge, whose content can in theory be quantified through Kolmogorov complexity.

Note also that none of the definitions prohibits that they can be linked to some system with certain physical or conceptual entities (make them represent something meaningful), or with metrics of values or utility (prioritize their importance). This linking is an additional question, which sometimes it is rather subjective, other times can be justified more objectively, but consists in a supplementary step. For example, equipped with a way to quantify information and knowledge, one could then ask the economically relevant question of how valuable a certain unit of information or knowledge is ($\frac{x \text{ value}}{\text{information bit}}$ or $\frac{x \text{ value}}{\text{algorithmic symbol}}$). Our foregoing definitions refer to the denominator of this more elaborated ratio and provide a way to unequivocally quantify plain information and knowledge, independent from its potential value or meaning.

In this sense our metrics are also more fundamental than the notions from economic and game theoretic decision theory (e.g. Gould 1974; Spence 1976; Hirshleifer and Riley, 1992), which defines information as the difference in payoff with and without information, as utility in a cost-benefit analysis, or as any other difference in value related to chance. This decision theoretic approach treats information as a black box by focusing on its effects (payoff, utility, cost, etc.), but not directly on the amount of information and knowledge itself. Later in this article we will link information to a specific value, its fitness value. This approach also provides a methodological bridge to economic decision theory (see Donaldson-Matasci, et al., 2010). However, metrics of information and knowledge are not automatically linked to value.

Tacitness: Economists are fascinated with the fact that not all the details of all knowledge structures are understood, articulable, or readily tractable. An agent might not be aware or might not be able to articulate the details of a procedure, which is then referred to as “tacit knowledge” (Polanyi, 1966). It is important to recognize that this does not change the fact that some kind of neural, social, or natural pattern underpins the executed procedure. At the end, some kind of result is repeatedly achieved through a certain (hidden) process. Information theorists speak about ‘hidden states’ and frequently model them with ‘hidden Markov models’, which imply that there are some states and dynamics that are not captured by the model. These are then modelled with probabilistic uncertainty (if they would be known, their deterministic algorithm would be revealed). In economic epistemology, the opposite of tacit knowledge is codified or explicit knowledge, whereas knowledge codification is defined as “the process of conversion of knowledge into messages which can be then processed as information” (Cowan and Foray, 1997; p.

596). In practice “it is difficult to argue that explicit knowledge, recorded in documents and information systems, is any more or less than information” (Rowley, 2007; p. 175). While this led to much confusion in economic literature (Cowan et al., 2000), from our information theoretic perspective the source of the confusion is clearly identifiable as the fact that both informational bits and algorithms consist of optimized symbols (often in binary form, such as 0s and 1s). However, our distinction consists in the fact that one is probabilistic (such as a chain of answers to questions), and the other deterministic (such as being an integral part of an explicitly codified procedure).

Heuristics: Behavioral economists often point out that some procedures are not optimal solutions for a specific task. They are merely approximate heuristics (Tversky and Kahneman, 1974). This also does not change the fact that some patterns does exist. Most practical computer algorithms work with shortcutting heuristics that provide an acceptable solution in return for speed and lower computational intensity.

The origin of knowledge: An important body of economic literature asks where new, innovative insights come from. The general consensus is that new knowledge emerges from the recombination of existing knowledge (e.g. Poincare, 1908; Schumpeter, 1939; Weitzman, 1998; Antonelli et al., 2010; Tria et al., 2014; Youn et al., 2015), be it by inductive observation or deductive inference. This notion is certainly very familiar to everybody who has written ‘new algorithms’ by ‘copy-pasting’ existing modules, subroutines, or callable function. While the notions here are certainly in agreement with these ideas, this article will not go further into this subject of the creation of new knowledge.

1.4 An application to growth economics

Applied to growth dynamics, the basic setup to illustrate the effect of information and knowledge on growth is similar to a fitness matrix such as customary in evolutionary game theory (e.g. Neumann and Morgenstern 1944; Nowak 2006). If the economic agent *knows* the unfolding dynamics of a deterministic environmental pattern, there is no doubt on how to best allocate resources. Table I shows a simplistic illustrative example with a two type population and the deterministic unfolding of two environmental states (“bull-“ and “bear-environment”). The two types could be two types of products, companies, or economic sectors, etc. It is customary in evolutionary theory to interpret the growth rate of a population as its fitness (biologists usually refer to the growth rate as the “rate of reproduction”, while game theorists refer to it as the “payoff”, Nowak, 2006; p. 14; 55). The fitness of each type in each environmental state is represented with the growth factor $w = \frac{\text{number of units at time } t+1}{\text{number of units at time } t}$. The units of growth can be US\$, number of workers, firms, patents, or any other countable variable or composite index. We use a capital letter W to indicate a superior growth rate over the inferior fitness: $W > w$. The table shows that the growth rate of type 1, W_1 , is larger than the growth rate of type 2, w_2 in a “bull-environment”, and the other way around for “bear-states”. If the agent is able to reallocate resources between types, this knowledge would allow the agent to optimize the population growth by always allocating 100 % of the resources at the share with the large fitness W_i , resulting in a long-term growth of: $\prod_t W_t = W_1^{\text{number of bull periods}} * W_2^{\text{number of bear periods}}$.

Knowledge about the unfolding of a dynamical system with two environmental states can be represented with a binary code (e.g. 1 and 0). The descriptive algorithm for Table I is: 110110110110... (bull-bull-bear –bull-bull -bear...). If this pattern is stationary in time, we can compress it by noticing that

a bull-state is always followed by two bear-states, resulting in a compressed algorithm that looks something like: [repeat 110]. This algorithm defines an endless time series in a deterministic manner. There would be no uncertainty in the future environmental pattern. The minimum number of symbols of the most efficiently compressed algorithm is its Kolmogorov complexity. Note that if there would be no discernable pattern and the sequence would be totally random, it would not be possible to compress the structure. We would need to record symbol by symbol, and the algorithm would be just as long as the sequence. The final length of the algorithm quantifies the amount of knowledge the agent needs to unequivocally describe such dynamical environmental system (which in the case of Table I is a periodic orbit). Since it describes a deterministic pattern in time, it is referred to as ‘knowledge’ here (knowledge about the unfolding environment). The agent ‘knows’ the environmental pattern.

Table I: Deterministic knowledge: illustrative setup of two types and two deterministic environmental states

Environmental state	Algorithm	Proportion of type 1: p_1	Proportion of type 2: $p_2 = (1-p_1)$
bull	1	W_1	w_2
bull	1	W_1	w_2
<i>bear</i>	0	w_1	W_2
bull	1	W_1	w_2
bull	1	W_1	w_2
<i>bear</i>	0	w_1	W_2
bull	1	W_1	w_2
bull	1	W_1	w_2
<i>bear</i>	0	w_1	W_2
bull	1	W_1	w_2
bull	1	W_1	w_2
<i>bear</i>	0	w_1	W_2
...

Shannon (1948) basically looked at the same issue the other way around. He asked: what is the likelihood that we pick one specific environmental state if we draw randomly from the universe of all possible states? Following the pattern from Table I, we can infer the probabilities of occurrence of each environmental state, with $e_1(bull) = \frac{2}{3}$ and $e_2(bear) = \frac{1}{3}$. Shannon derived with metric of information (entropy) from the uncertainty contained in a distribution like this. He concluded that information is the amount of uncertainty that is reduced when we are ‘informed’ about the correct state (the amount of surprise of finding to be in one or the other state). Table II presents a probabilistic version of the previous setup from Table I. Again, type 1 is better adapted to the environment $e_1(bull)$, and type 2 to environment $e_2(bear)$. Here we only have probabilistic and partial insight into the unfolding of the dynamical system of the environment. We know the probabilities of the occurrence of an environmental state (its probability of occurrence), but lack insight into the temporal sequence of the events. The uncertainty metric entropy evaluates the uniformity of the random variable. It reaches its maximum at a

uniform distribution (in the binary case [0.5,0.5]) and its minimum in the case in which only one realization of the random variable has all probability ([1;0] or [0;1]) (Cover and Thomas, 2006).

While it is straightforward to see that deterministic knowledge about the environment optimizes growth by reallocating 100% of the resources to the type with the larger fitness W in each period (Table I), it is not as intuitive to see how to optimize the total population fitness in the case of mere partial probabilistic information about the unfolding of the environment (Table II). This questions lead us to an important result from information theory. As far back as 1956, Kelly had shown that if (a) populations / sectors are grouped in a way that different types specialize on different environments; (b) resources can be reallocated from one type to the other; and (c) at least partial information can be extracted about a future state of the environment (i.e. a probability distribution is known), growth can be optimized by adjusting and maintaining the proportions of the evolving population (Kelly, 1956). The resulting optimal growth rate is higher than the growth rate achieved by blind natural selection can be achieved (for an overview see Cover and Thomas, 2006; Ch.6). The beauty of the result consists in the fact that the achievable increase in fitness is equivalent to the amount of information about the future state of the environment (the mutual information $I(E; C)$), which in the case of perfect information, converges to deterministic knowledge about the pattern of the environment.

Table II: Probabilistic information: illustrative setup of two types and two probabilistic environmental states

Probability of environmental state	Proportion of type 1: p_1	Proportion of type 2: $p_2 = (1-p_1)$
e_1 (bull)	w_1	w_2
e_2 (bear)	w_1	w_2

2. Informed intervention

The idea of bet-hedging for risk reduction under uncertainty (also known as stochastic switching) goes at least back to Bernoulli in 1738 (Stearns, 2000). Kelly's result from 1956 formalized important aspects of it and the notion of bet-hedging also found its way into financial portfolio theory (Latané, 1959), where it became a cornerstone of successful long term investment strategies, such as Warren Buffet's (Pabrai, 2007). Important work has recently be done by applying it to biological evolution (Bergstrom and Lachmann; 2004; Kussell and Leibler, 2005; Donaldson-Matasci, Bergstrom and Lachmann; 2010; Rivoire and Leibler, 2011). The setup links probabilities of occurrence and shares of proportions with the values of a fitness matrix (such as in Table II). As such, it does link information to value (the respective growth factor w). As a result, information theoretic bet-hedging can closely be linked to decision theory of behavioral economics (e.g. Gould, 1974; for a joint treatment of the biological and behavioral economics approaches see Donaldson-Matasci, et al., 2010). Similar results have been derived (seemingly independently) by Blume and Easley (1993, 2002) and led to considerable research in the evolutionary finance (e.g. see special issue of the Journal of Mathematical Economics, 41 (1-2); Hens and Schenk-

Hoppe, 2005). While most of these results derive similar conclusions, especially those from economics literature do not use the mathematical formulations of information theory. Here we will therefore follow Kelly's original logic from 1956 (and its direct extensions) to clarify the fundamental role of information and knowledge in structural change.

2.1 Kelly's bet-hedging

The basic idea behind Kelly's bet-hedging (also called Kelly gambling, or the Kelly criterion) is to not maximize short-term utility, but to consider the long term return over time. To optimize the long-term growth rate, we have to adjust our proportions of types (which refer to a variety in "space") to the probabilities of environmental states (which refer to a variety in "time"). Kelly (1956) showed that in the extreme case where the inferior fitness of a given environmental state equals 0 (in our example: $w_1 = 0 = w_2$, resulting in a diagonal fitness matrix with non-zero values only in the diagonal), the best we can do is to maintain the shares of types proportional to the occurrence of their respectively favorable environments. A diagonal fitness matrix implies a 'winner-takes-it-all' logic, where only only highly specialized type survives. Following our case of Table II, this would result in $e_1(bull) = p_1 = \frac{1}{3}$ and $e_2(bear) = p_2 = \frac{2}{3}$ (see Cover and Thomas, 2006; p. 161). The technical reason behind the superiority of proportional betting is that a geometric expansion always overtakes arithmetic expansion (Cover and Thomas, 2006, Ch. 6), which is a well-known result that will not have to be explained to readers of this Journal. The superiority of geometric expansion (or compound interest) has become second nature to most economists. The result is a clear-cut one-to-one relationship between what is likely to be expected ('in time') and the allocation of current resources right now ('in space').

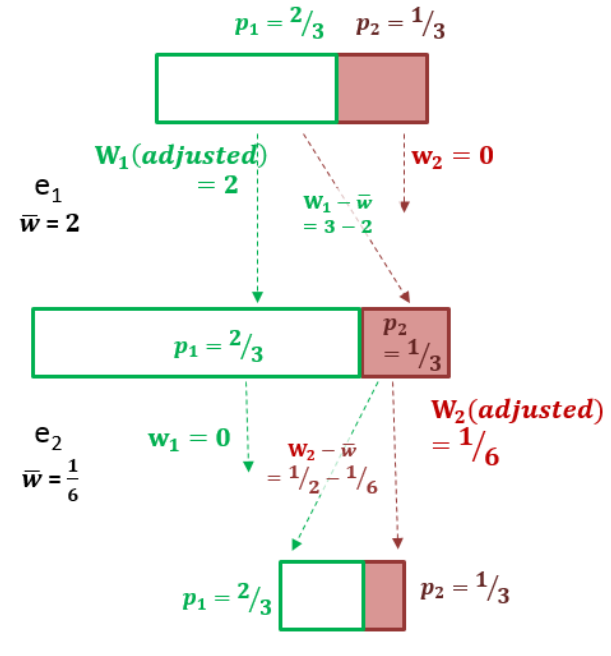
Since differences in the fitness of types destroys proportionality through evolutionary selection (whereas the fitter gains share), a proportional betting strategy requires a constant redistribution of resources between types to maintain proportionality. This requires taking resources from the faster growing types and reallocating them them to the less fit types to maintain proportionality.¹ Maintaining proportionality implies that this taking and giving adjusts the fitness of each type to the average population fitness, which is defined as the mean population fitness $\bar{w} = \sum_i p_i * w_i = E[w_i]$ (the expected value over the fitness of all types) (see Figure 1 for a schematic example).²

Figure I: Bet-hedging implies a constant redistribution of resources to maintain stable proportions of types, stopping evolutionary (Darwinian) selection in favor of informed (Lamarckian) intervention

¹ The exception being independence on the lower level of subtypes. If the distribution of the types of the lower level is independent from the distribution of the types on the higher level ($p_i \cap p_{ij} = p_i * p_{ij}$), the lower level types will create the same growth rates for all higher level types, which will naturally maintain a unchanging and stable distribution on the higher level (maintain proportionality).

² If we want to maintain $p_i = p_i^{t+1}$, we need to only pass part of the faster growing type on to the next generation ($p_i * \frac{\bar{w}}{w_i} * \frac{w_i}{\bar{w}} = p_i^{t+1}$), and redistribute the rest to the less growing type, which is the remaining $(1 - \frac{\bar{w}}{w_i})$.

	$p_1 = 2/3$	$p_2 = 1/3$	Population Fitness
$e_1 = 2/3$	$W_1 = 3$	$w_2 = 0$	$\bar{w}_1 = 2$
$e_2 = 1/3$	$w_1 = 0$	$W_2 = 1/2$	$\bar{w}_2 = 1/6$



The long term growth rate is calculated by multiplying the generational growth rates of the population: $\overline{W}^{long\ term} = \overline{w}^t * \overline{w}^{t+1} * \overline{w}^{t+2} * \dots$ Applied to our binary state environment from Table II, and sticking to the extreme case in which only the specialized type survives in a specific environment (minuscular $w = 0$, see Figure I), and defining that $w_{i|e_1}$ corresponds to the fitness given environment e_1 , the population growth over a typical sequence of N periods is:

$$\begin{aligned}\overline{W}^{long\ term} &= (E[w_{i|e_1}])^{N * e_1} * (E[w_{i|e_2}])^{N * e_2} = \text{with } (w_i = 0) \\ &= (p_1 * W_1)^{N * e_1} * (p_2 * W_2)^{N * e_2}\end{aligned}\quad (1)$$

Maximizing the long term growth rate is equivalent to maximizing the generational exponent of growth, which suggests taking the logarithm and divide by N (assuming a large number of N):

$$\begin{aligned}\rightarrow \frac{1}{N} \log \left[\prod_i (p_i * W_i)^{N * e_i} \right] &= \log[\overline{W}] = \sum_i e_i \log(p_i * W_i) \\ &= \sum_i e_i \log(W_i) + \sum_i e_i \log(p_i)\end{aligned}\quad (2)$$

$$\begin{aligned}&= \sum_i e_i \log(W_i) + \left[\sum_i e_i \log(e_i) - \sum_i e_i \log(e_i) \right] + \sum_i e_i \log(p_i) \\ &= \sum_i e_i \log(W_i) + \sum_i e_i \log(e_i) - \sum_i e_i \log\left(\frac{e_i}{p_i}\right) \\ &= \sum_i e_i \log(W_i) - H(E) - D_{KL}(E||P)\end{aligned}\quad (3)$$

The entire derivation is presented so the reader can appreciate the simplicity of the algebraic reformulations that result in the appearance of the information theoretic metrics H and D_{KL} . Equation (3) shows that the logarithmic time average of the population fitness $\log[\overline{W}]$ consists of three components. The first sets the benchmark of optimal fitness: $\log[\overline{W}] = \sum_i e_i \log(W_i)$. This implies the allocation of all resources in a specific environmental state to the type with superior fitness case. Reverting the logarithm shows this clearly: $\overline{W} = \prod (W_i^{e_i})$ (compare with the case of deterministic knowledge about the environmental state in Table I). Furthermore, equation (3) shows that the attainability of this optimal growth factor depends on two other components. First, it depends on the uncertainty of the environment, its entropy $H(E)$. Entropy is one of the most fundamental measures in all science (Shannon, 1948; Cover and Thomas, 2006). Large uncertainty implies more uniformly distributed states of the environment and therefore high entropy $H(E)$. This reduces the long term average growth rate (for our binary case, the maximum uncertainty occurs at the 50%-50% split, which accounts for 1 bit of environmental uncertainty). If there is no uncertainty, the entropy of the environment becomes 0, and we are left with the case of 'complete information' or a 'perfect cue', which enables us to pick type 1 in environment 1, and type 2 for environment 2. The result will be equivalent with a deterministic resource allocation strategy, which can be quantified by Kolmogorov's metric for algorithmic knowledge (in the sense of Table I).

Secondly, the attainable growth factor depends on the divergence between the population distribution and the distribution of the environment. This is measured by the relative Kullback-Leibler entropy $D_{KL}(E||P)$ (Kullback-Leibler, 1951), which quantifies the informational inefficiency to assume

that the distribution is P , when the true distribution is E . It is another one of the fundamental metrics of information theory and just like absolute entropy, relative entropy is also always positive. It quantifies the familiar notion that fitness depends on the match between the population and the environment. The better the fit, the higher fitness: “fit-ness”, in the true sense of the word. Since the logarithm is a monotonically increasing function, the larger the average ratio of $\left(\frac{e_i}{p_i}\right)$, the larger the Kullback-Leibler information between our betting strategy and the environment, and the more reduced our average long-term growth rate. Proportional betting (with $p_i = e_i$) therefore achieves the highest attainable long term population fitness by eliminating this term: $\log\left(\frac{e_i}{p_i}\right) = \log(1) = 0$. This implies that the time average of the population fitness is only reduced by the uncertainty of the environment (quantified with Shannon’s entropy, whereas E_i stands for the expected values over all states):

$$\log[\bar{W}] = E_i[\log(W_i)] - H(E) \quad (4)$$

This provides a formal way to quantify the long-standing intuition of economists that considers knowledge “a form of adaptation of human beings to the external environment in which they live” (Saviotti, 2004; p. 101). If deterministic knowledge provides the probability distribution of the environment (in form of a probability ‘lookup table’), the proportions of resources can be adjusted to be perfectly adapted to the environment. Our equation shows that this will also optimize the growth potential. Optimal adaption means optimal fit and implies optimal fitness.

2.2 The fitness value of information

We can also reduce uncertainty with additional side information about the future state of the environment. The two sources for possible side information can be the observation of the past (‘memory’) and observations of third events from the present that correlate with the future (‘cues’). Let us start with supposing that we have a certain amount of memory that allows us to extract information about the future. The importance of data analysis from past patterns to make predictions becomes increasingly important in a big data world (Hilbert, 2016). We will ask how much such information from memory can increase our maximal growth rate. The memory about the past represents a new conditioning variable C . Based on one kind of past or another (on the realizations of the random variable C), we obtain a conditional likelihood of the occurrence of a certain environmental state. For example, the past could have been a prolonged bull market. The probabilities of a future bull or bear market conditioned on this past are conditional probabilities $e(\text{bull}|c = \text{bull})$ and $e(\text{bear}|c = \text{bull})$. We therefore end up with two separate environmental distributions, which are distinguished by the conditional variable C . This will result in two different proportional bet-hedging distributions to optimize growth, depending on the realization of the conditioning cue.

We start with equation (2) from above and adjust for the fact that we now work with the fine-tuned conditional probability $e(i|c)$ for the environment and $p(i|c)$ for the respective proportional bets, and adjust for how often the bull or bear past occurs with $e(c)$. Aiming at maximal growth rates we stick to proportional betting, turning the D_{KL} term of equation (3) to 0. Therefore, the long term optimal growth rate, conditioned on this memory is:

$$\log[W^{opt.}(E|C)] = \sum_c e(c) \sum_i e(i|c) [\log(W_i) + \log(p(i|c))] \quad (5)$$

$$= \sum_{c,i} e(i,c) \log(W_i) + \sum_{c,i} e(i,c) \log(p(i|c)) = \text{for proportional betting } (p = e)$$

$$= E_i[\log(W_i)] - H(E|C) \quad (6)$$

The resulting term is the conditional entropy $H(E|C)$. It quantifies the amount of uncertainty that is left after the conditional variable is known (Cover and Thomas, 2006). For the difference between the optimal long term growth rate with and without memory (subtracting equation (4) from equation (6)) we get:

$$\Delta W^{optimized} = \log[W^{opt.}(E|C)] - \log[W^{opt.}(E)] =$$

$$= \left[\sum_i e(i) \log(W_i) - H(E|C) \right] - \left[\sum_i e(i) \log(W_i) - H(E) \right] = H(E) - H(E|C)$$

$$= I(E; C) \quad (7) \text{ Fitness value of information}$$

This measure is Shannon's celebrated mutual information (Shannon, 1948). In our case it is the mutual information between the informational cue extracted from the past, and the current environment. If the logarithm to base 2 is used, it is measured in bits. It quantifies how much the memory of the past can tell us on average about the future (this relation is of course noisy and therefore probabilistic). In words, it says that information about the past can increase the optimal growth rate exactly by as much as the past can tell us about the future. If the past tells us a lot about the future, the mutual information between past and future will be high, and fitness can be increased decisively. If the future is independent from the past, the mutual information is 0 (Cover and Thomas, 2006), and no increase is possible. This one-to-one relationship between the amount of information (as quantified by Shannon) and the increase in growth rate provides information with a value, a fitness value.

The fact that Shannon's probabilistic information metric arises in these equation can be intuitively understood as follows (Bergstrom and Lachmann; 2005): imagine complete uncertainty among 6 equally likely choices (e.g. investment options). Having US\$ 6, the bet hedging investor would distribute US\$1 to each choice and surely walk out with one winning dollar. Let us assume that we receive 1 bit of information, which is defined by Shannon (1948) as the reduction of uncertainty by half. The investor would now know which half of the investment options fail. He would now equally distribute US\$2 of his US\$6 to each of the remaining 3 choices and walk out with two winning dollars. The reduction of uncertainty increased the growth of the payoff. The equation shows that the logic of dividing uncertainty by 2 will increase gains by 2 is not a numerical coincidence, but a solid relation between uncertainty reduction and achievable growth.

We can also include both the past and a related cue (e.g. the behavior of other economic agents, or obtain a tip from an expert), and work with two conditioning factors: both cases (with and without cue) are conditioned on a series of past environmental states, but the case with cue is now also conditioned on a series of cues from different points in time of the past (receiving one cue per generation). It is important to notice that cues from the past can still tell us something about the upcoming present (additionally to the present cue). Permuter, Young-Han and Weissman (2011) have shown that instead of Shannon's mutual information, the correct measure in this more general case is Massey's directed

information $I(\text{cues}^n \rightarrow T^n)$ (Massey, 1990; Massey and Massey, 2005; also Kramer, 1998). If the present environmental state and the side information are pairwise independent (e.g. the past cue does not tell us anything about the present environment), then the Massey's directed information becomes Shannon's mutual information, which coincides with Kelly's original result (1956).

In short, the information from past experience and from present cues can allow us to predict the future and allows us to increase the growth rate of the population if we allocate resources accordingly. This result stems from the theorem that conditioning reduces entropy/uncertainty ("information can't hurt"; Cover and Thomas, 2006; p.29).³

2.3 Extension to mixed fitness landscapes

Kelly's classical result provides the most important intuition of how information is linked to growth. However, the math works out as nicely because it is based on the special case of a diagonal fitness matrix, with non-diagonal values being 0 and cancelling out (see Figure I and equation (1)). Over recent years, the result has been expanded to any kind of mixed (non-diagonal) fitness matrix (see Donaldson-Matasci, et al., 2008; 2010; Rivoire and Leibler, 2011; Hilbert, 2015). The mathematics behind this generalization represents any non-fatal fitness value as a combination of fitness values for a hypothetical diagonal fitness matrix. In other words, it assumes a hypothetical world with one perfectly specialized type per environment and proposes that any existing type fitness is a combination of those specialized fitness values over the different environmental states. The conditional weighting matrix $E|M$ expresses the mixed growth landscape in which each environmental state has a specialized type. The matrix is stochastic (conditioned, with columns summing up to 1 over all environmental states), as it weights each type over the different environmental states. It answers the question: if one mixed type would be a combination of totally specialized types, how would they be weighted? Therefore, the reconstruction of the empirical mixed fitness matrix in terms of a weighted idealized fitness matrix can be expressed in terms of matrix algebra.

$$\begin{bmatrix} W_1 & w_2 \\ w_1 & W_2 \end{bmatrix} = \begin{bmatrix} W_1^{hyp} & 0 \\ 0 & W_2^{hyp} \end{bmatrix} * \begin{bmatrix} e_1|m_1 & e_1|m_2 \\ e_2|m_1 & e_2|m_2 \end{bmatrix} \quad (8)$$

Solving for these two unknown matrixes is straightforward and can almost always be done uniquely⁴. We now have the mixed matrix in terms of a diagonal matrix and Kelly's result tells us how to optimize diagonal matrixes: through a one-to-one proportionality with the environment. This suggests that the mixed fitness landscape matrix can be optimized by optimizing the diagonal matrix. Therefore, we optimize the diagonal matrix by weighing it with the environmental weights e_i and solve for the optimal distribution of types p_i for the mixed matrix.

$$\begin{bmatrix} W_1 & w_2 \\ w_1 & W_2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} W_1^{hyp} & 0 \\ 0 & W_2^{hyp} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (9)$$

³ Conditioning reduces uncertainty only on average. A particular realization of the conditioning variable might increase uncertainty (for example new piece of evidence for an investor), but on average (e.g. considering the evidence and its complement) uncertainty is reduced.

⁴ Note that Donaldson-Matasci, et al. (2008, 2010) flip the weighting matrix, with rows summing up to 1.

Solving for the optimal distribution of types P most often results in a distorted proportionality between the distribution of the environment and the growth-optimal distribution of types (depending on the combination of growth rates in the matrix of Table II). The nice one-to-one proportionality between the distribution of types and environmental states is distorted. Additionally, unfortunately, the math works out in such a way that some combination require a distribution of types that call for $p_i < 0$ or $p_i > 1$, which is of course undefined. This suggests to bet all resources on the share with $p_i > 1$ and none on the share with $p_i < 0$, which has been called a “pure strategy” (Rivoire and Leibler, 2011), i.e. all bets on one type. Between those extremes exists the so-called “region of bet-hedging” (Donaldson-Matasci, et al., 2010) a in which bet-hedging suggests for a “mixed strategy” with certain proportions of types.

Hilbert (2015) derived a complete decomposition of evolutionary fitness in information theoretic terms, which shows that the constraint from the mixed fitness landscape is also a relative entropy. It quantified the informational divergence between the the environment from the point of view of the weighting matrix, and from the point of view from the population after average updating at time $t +$: P^{t+} . When growth is optimized, this results in a form that corresponds to Kelly’s equation (see equation (3)):

$$\log(\bar{W}) = \underbrace{E[\log^d W]}_{\text{noiseless channel}} - \underbrace{H(E)}_{\text{env.uncertainty}} - \underbrace{D_{KL}(e|p^{t+} || e|m)}_{\text{empirical landscpae constraint}} \quad (10)$$

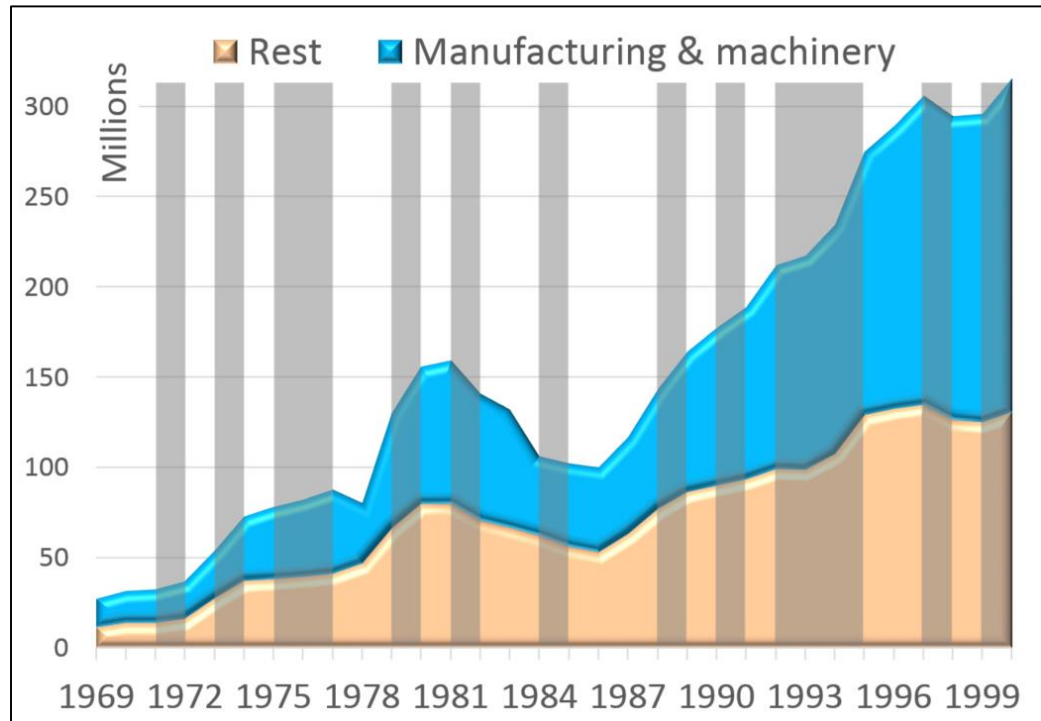
This last D_{KL} terms disappears in the cases where bet-hedging is possible (be it one-to-one proportional for Kelly’s Kelly’s diagonal fitness matrix, or non-proportional bet-hedging for mixed fitness landscapes ‘within the region of bet-hedging’). This takes the same form as equation (4). Outside the region of bet-hedging (when all resources are best allocated on one specific type, while the other type receive no resources) this D_{KL} term reduces the achievable level of fitness.

3. Empirical application

Let us explore an empirical application of the presented logic to illustrate the role of information in structural change. We use export data from UN COMTRADE (UNSD, 2012), more specifically, the cleaned NBER version of the dataset from Feenstra, et al. (2005) (SITC, rev.2).⁵ Our population structure distinguishes between manufacturing and non-manufacturing sectors in U.S. exports (i.e. grouping the first digit sectors “6 Manufactured goods classified chiefly by materials”, “7 Machinery and transport equipment”, “8 Miscellaneous manufactured articles” versus the remaining sectors of digits 1-5). Over the three decades, the manufacturing sector varies between 41.3 % and 58.6 % of the total, while U.S. export grows from US\$ 27.2 to 316.61 billion, with an annual compound growth rate of 8.227 % (see Figure II).

⁵ Since the records are often still incomplete (which introduces disruptive discontinuity in our time-sensitive evolutionary analysis), we limit the dataset to the 512 four digit export items with the most complete record. On average, roughly 1,300 four digit items are presented in the database, but many of them contain records only for a couple of sporadic years. The cutoff of 512 was defined because otherwise items that have less than 50% of its time history record complete would have to be included into the analysis (sporadically missing records were interpolated).

Figure II: Evolution of U.S. export structure (1969-2000) based on the 512 most complete export fourth digit SITC rev. 2 items, as shares of nominal millions US\$.



Source: based on UNSD (2012) and Feenstra, et al. (2005).

We choose to detect an environmental distribution E by counting the shares in which one subsector grows faster than the other one. It turns out that 64.5 % of the years are a manufacturing friendly environment. This allows us to empirically detect the corresponding fitness matrix

$$\begin{array}{cc}
 p(\text{rest}) & p(\text{manuf}) \\
 e(\text{rest}) & 1.135 & 1.004 \\
 e(\text{manuf}) & 1.054 & 1.128
 \end{array}
 \text{ . Equation (8) allows to calculate an empirical version of equation (9):}$$

$$\begin{bmatrix} 1.135 & 1.004 \\ 1.054 & 1.128 \end{bmatrix} \begin{bmatrix} p(\text{rest}) \\ p(\text{manuf}) \end{bmatrix} = \begin{bmatrix} 2.978 & 0 \\ 0 & 1.702 \end{bmatrix} \begin{bmatrix} 0.355 \\ 0.645 \end{bmatrix}$$

Solving it shows that the result lies within 'region of bet-hedging' and that optimality is achieved if the manufacturing sector would represent a stable share of $p(\text{manuf}) = 0.598$ of the export. Following this strategy would increase total export to US\$ 325.68 billion or a compound annual fitness rate of 8.337 %.

An ingenious economist or econometrician could now look for side information linked to this evolutionary trajectory. Imagine some kind of big data analysis concludes that an important indicator is the exports of machinery and transport equipment of neighboring Canada (taken from the same statistical source). We could therefore devise a different strategy if Canadian machinery export of the previous year has increased (these years are indicated with grey bars in Figure II). This introduces a conditioning variable, and results in two different strategies. The application of equations (5) and (6) recommends to opt for a

stable share of $p(\text{manuf}) = 54.6\%$ if last year's Canadian machinery exports have increased, and a share of $p(\text{manuf}) = 64.2\%$ if not. This fine-tuned strategy increases total exports to US\$325.78 or a compound annual growth rate of 8.338. While this might not seem as much, the driver of this increase in growth has been plain information. In terms of equation (7), this can be shown quantitatively (entropy measured in information bits with a logarithm of base 2):

$$\begin{aligned} \log[W^{\text{opt.}}(E|C)] - \log[W^{\text{opt.}}(E)] &= \log[1.08338] - \log[1.08337] = H(E) - H(E|C) \\ &= 0.93831 - 0.93830 = 0.00001 = I(E; C) \end{aligned}$$

The achieved increase in fitness is exactly equal to the mutual information between the cue (Canada's machine export) and the environmental pattern of the U.S. export. There is not much mutual information between those two, but even this can make a difference, as it would increase U.S. exports by US\$ 100. Of course, finding strong predictors are at the heard of much econometric and machine learning analytics.

A perfect cue would imply complete deterministic knowledge of the entire environmental pattern (in line with Table I). A complete algorithm would tell us with unequivocal foresight which sector would grow faster in which year, and in theory allow us to allocate all available resources to this sector to exploit this higher growth to the fullest. Following our empirically identified fitness matrix and its environmental distribution, this would give us $[1.135^{0.355} * 1.128^{0.645} = 1.1306]$, or a compound annual growth rate of 13.06 %. This implies that the fitness value of complete knowledge would be an additional 4.8 % of annual growth (compared to the empirically detected growth rate), which would catapult U.S. exports to US\$ 1.22 trillion over 31 years.

4. Discussion of limitations and extensions

This final section emphasizes the implicit assumptions of the presented approach, outlines some of its limitations, and presents possible extensions and outstanding challenges for a future research agenda.

Exogenous growth: The presented approach is an exogenous growth model. It does not explain where the fitness landscape of Tables I and II come from. They might be internally produced or externally given, but they assumed to be found in empirical reality.

Environmental stationarity: Probabilistic bet-hedging requires the fitness landscape to be stationary. Otherwise the environmental distribution could not be calculated. Cherkashin, Farmer and Lloyd (2009) analyzed what happens in the case of density dependence, that is when the odds of environmental occurrence E is influenced by the allocation of resources P . This feedback loop endogenizes the environmental distribution. They show numerically and analytically that wealth converges on a stable environmental distribution and a single optimal strategy. In general, if types foster unfavorable environments, mixed bet-hedging are optimal, while pure extremist strategies are superior for types that fosters favorable environments.

Easy resource redistribution: Bet-hedging assumes that one can readily redistribute resources between types. This might not work if resource investments are irreversible and bound for a long time (Demers, 1991), have some lag-time (even an investment firm cannot move several billion dollars immediately through a market), or are not readily convertible (e.g. human capital).

Darwin vs. Lamarck: Our analysis does not define the practical nature of this redistribution mechanism, nor how it arises. In biological populations, bet-hedging (also called ‘stochastic switching’) is often the result of blind Darwinian mutations (e.g. Maamar, Raj and Dubnau, 2007). What is born out of an accident, is adopted through the resulting superior growth rate. In the economy, information can be used proactively to guide Lamarckian intervention into the system, executed through a large variety of resource redistribution mechanisms, including taxes and government expenditures, insurances, private sector cross-subsidies, trade and capital restrictions, and others.

Squared Fitness Matrix: We worked with an equal number of types and environmental states (of arbitrary fitness matrix size $n \times n$). This assumption captures Ashby’s (1956) law of requisite variety. The model naturally becomes more complex as the subdivision of type specializations becomes multiary (which requires the definition of redistribution among more than two types) or when the number of types does not correspond to the number of environmental states.

Information asymmetry: Another assumption is that all actor types perceive the same informational cues from the environment. Rivoire and Leibler (2011) show that a randomization of cues actually increases the achievable growth rate, since the value of information acquired collectively by the heterogeneous population exceeds the value of the information acquired by any homogenous group of types with identical perception. This is in agreement with Hayek’s (1945) classical argument that a centrally planned market could never match the efficiency of the decentralized market because any individual knows only a small fraction of all which is known collectively.

Information cost: Our setup no cost for information perception, collection and analysis. This means that cues and memory come for free. It is straightforward to include a cost variable and this will obviously reduce the lucrativeness of bet-hedging. Kussell and Leibler (2005) showed that a population can afford to pay more for information the more uncertain its environment.

Average results: There are two mathematical caveats for working with averages in space and time. First, conditioning reduces uncertainty and creates information only on average.³ A specific piece of evidence can rather create confusion and increase uncertainty, which can lead to challenges in particular real-world cases. Second, a typical limitation of results based on geometric expansion is the fact that the exact result only holds for long periods of time (in the limit asymptotically). This is in principle bad news, since asymptotic limits can take too long as to be relevant in practice. The good news is that Cherkashin (2004) was able to show that it also holds for finite period provided that their duration is not known to the types, which is a more realistic assumption.

Evolution of what? A final fundamental question refers to how to define an evolving population? In our export example we distinguished between manufacturing and non-manufacturing products. But who assures us the the evolving economy cares about these two sectors? The chosen taxonomy affects the identified fitness landscape, its stationarity and the involved amount of information processing. We could as well differentiate among organizational forms as our units (Aldrich and Ruef, 2006), or classify types according to routines (Nelson and Winter, 1985) or cultural norms and institutionalized habits (Boyd and Richerson, 2005). Much future research is required to understand the fundamental units with which the economy processes information. The presented approach gives a quantitative formal way to go about it.

5. Conclusion: the quantifiable role of information for growth

Returning to our example from the introduction, conditioned on past memory or a cue for the binary random variable rain/sunshine, a bakery can calculate likelihoods and therefore proportional shares for the optimal structural portfolio for salty and sweet goods. This leads to more fine-tuned (conditioned) strategies (e.g. one strategy conditioned on the realization of rain, and another one for sunshine). The potential increase in growth depends on the amount of mutual information between the random variable 'weather' and the sales pattern of baked goods. Mathematically, this increase is due to the fact that conditioning reduces entropy/uncertainty³ (equation (7)). In the case of a 'perfect cue', this string of information turns into 'deterministic knowledge' of the environmental pattern, which allows to flexibly adjust the structural portfolio to each environmental condition, optimizing achievable growth. In this case we 'know' the deterministic future pattern (knowledge). Both, the amount of information extracted through probabilistic information and the amount of information contained in a deterministic algorithm are asymptotically equivalent. Just like Maxwell's famous demon (Zurek, 1989) has two choices to extract energy (the potential to do work) from its environment (either observe or know), we have seen that the potential to grow social systems also hinges on either information or knowledge. The question of how structural change can be used as a tool to obtain growth potential depends on the amount of information that is obtainable from environmental patterns. If we have deterministic knowledge about the future, we could optimize the economic structure accordingly. With probabilistic information, we still can, but in a limited fashion.

This leads to the final question of how much information and knowledge is obtainable about the short- and long term patterns. In essence, this question is at the heart of what economic analysis does: identifying patterns in market forces and human behavior in order to extract laws and rules of thumb that allow us describe and predict the dynamics of the system and act accordingly. In this sense, the presented approach shows that economics itself is a quantifiable input ingredient of economic growth. The more we know about the economy and its dynamics, the more we can allocate resources to enable growth. Information is not sufficient, but it is the necessary ingredient for growth. We can and should start to quantify its crucial role.

Acknowledgements

I am indebted with Matina Donaldson-Matasci, Steve Frank, Pier Pablo Saviotti, Cristiano Antonelli, Peter Monge, Wojciech Zurek, Carlton Caves, Srideep Musuvathy, William Wu, Gerhard Kramer, Manuel Castells, Bill McKelvey, United Nations ECLAC, and the faculty of the Santa Fe Institute, who were all instrumental in the formation of my opinion about these issues over recent years and patiently answered all my questions in personal conversations.

References

- Ackoff, R. L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis*, (16), 3–9.
- Aldrich, H., & Ruef, M. (2006). *Organizations Evolving* (2nd ed.). Sage Publications Ltd.
- Antonelli, C., Krafft, J., & Quatraro, F. (2010). Recombinant knowledge and growth: The case of ICTs. *Structural Change and Economic Dynamics*, 21(1), 50–69. <http://doi.org/10.1016/j.strueco.2009.12.001>
- Ashby, W. R. (1956). *An introduction to cybernetics*. London: Chapman & Hall. Retrieved from <http://pcp.vub.ac.be/books/Introcyb.pdf>
- Bateson, G. (1972). *Steps to an Ecology of Mind*. Random House, New York.
- Becker, S. C. (2003). The concept of routines: some clarifications. *Cambridge Journal of Economics*, 29(2), 249–262.
- Bennett, C. H. (1982). The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21, 905–940.
- Bennett, C. H., Gacs, P., Li, M., Vitanyi, P. M. B., & Zurek, W. H. (1998). Information distance. *IEEE Transactions on Information Theory*, 44(4), 1407–1423. <http://doi.org/10.1109/18.681318>
- Bergstrom, C. T., & Lachmann, M. (2004). Shannon information and biological fitness (pp. 50–54). Proceedings of the IEEE Workshop on Information Theory. Retrieved from <http://octavia.zoology.washington.edu/publications/BergstromAndLachmann04.pdf>
- Bergstrom, C. T., & Lachmann, M. (2005). The fitness value of information. *arXiv:q-bio/0510007*. Retrieved from <http://arxiv.org/abs/q-bio/0510007>
- Bernstein, J. H. (2011). The Data-Information-Knowledge-Wisdom Hierarchy and its Antithesis. *NASKO*, 2(1), 68–75. <http://doi.org/10.7152/nasko.v2i1.12806>
- Blume, L. E., & Easley, D. (1993). Economic natural selection. *Economics Letters*, 42(2-3), 281–289.
- Blume, L. E., & Easley, D. (2002). Optimality and Natural Selection in Markets. *Journal of Economic Theory*, 107, 95–135. <http://doi.org/10.1006/jeth.2000.2764>
- Boyd, R., & Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. Oxford University Press, USA.
- Brookshear, J. G. (2009). *Computer Science: An Overview* (10th ed.). Addison Wesley.
- Caves, C. (1990). Entropy and Information: How much information is needed to assign a probability? In W. H. Zurek (Ed.), *Complexity, Entropy and the Physics of Information* (pp. 91–115). Oxford: Westview Press.
- Chaitin, G. J. (1966). On the Length of Programs for Computing Finite Binary Sequences. *Journal of the ACM (JACM)*, 13, 547–569. <http://doi.org/http://doi.acm.org.libproxy.usc.edu/10.1145/321356.321363>
- Cherkashin, D. (2004). *Perception game* (Ph.D. thesis). University of Chicago, Dept. of Mathematics.
- Cherkashin, D., Farmer, J. D., & Lloyd, S. (2009). The reality game. *Journal of Economic Dynamics & Control*, 33(5), 1091–1105.
- Christensen, B. (2012, April 24). Smarter Analytics: Der Bäcker und das Wetter [the baker and the weather] [IBM Analytics]. Retrieved from <https://www.youtube.com/watch?v=dj5iWD2TVcM>
- Cohendet, P., & Steinmueller, W. (2000). The Codification of Knowledge: A Conceptual and Empirical Exploration. *Industrial and Corporate Change*, 9(2), 195–209.
- Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd Edition). Hoboken, NJ: Wiley-Interscience.
- Cowan, R., David, P. A., & Foray, D. (2000). The explicit economics of knowledge codification and tacitness. *Industrial and Corporate Change*, 9(2), 211–253. <http://doi.org/10.1093/icc/9.2.211>
- Cowan, R., & Foray, D. (1997). The Economics of Codification and the Diffusion of Knowledge. *Industrial and Corporate Change*, 6(3), 595–622. <http://doi.org/10.1093/icc/6.3.595>
- Crutchfield, J., & Feldman, D. (2003). Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 13(1), 25–54.

- Crutchfield, J. P. (2012). Between order and chaos. *Nature Physics*, 8(1), 17–24.
<http://doi.org/10.1038/nphys2190>
- Demers, M. (1991). Investment under Uncertainty, Irreversibility and the Arrival of Information Over Time. *The Review of Economic Studies*, 58(2), 333–350. <http://doi.org/10.2307/2297971>
- Donaldson-Matasci, M. C., Bergstrom, C. T., & Lachmann, M. (2010). The fitness value of information. *Oikos*, 119(2), 219–230. <http://doi.org/10.1111/j.1600-0706.2009.17781.x>
- Donaldson-Matasci, M., Lachmann, M., & Bergstrom, C. T. (2008). Phenotypic diversity as an adaptation to environmental uncertainty. *Evolutionary Ecology Research*, 10, 493–515.
- European Commission. (2014). Communication on data-driven economy. Retrieved April 27, 2015, from ec.europa.eu/digital-agenda/en/news/communication-data-driven-economy
- Feenstra, R. C., Lipsey, R. E., Deng, H., Ma, A. C., & Mo, H. (2005). *World Trade Flows: 1962-2000* (Working Paper No. 11040). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w11040>
- Gell-Mann, M. (1995). *The Quark and the Jaguar: Adventures in the Simple and the Complex*. New York: St. Martin's Griffin.
- Gould, J. P. (1974). Risk, stochastic preference, and the value of information. *Journal of Economic Theory*, 8(1), 64–84.
- Hayek, F. A. (1945). The Use of Knowledge in Society. *The American Economic Review*, 35(4), 519–530.
- Hens, T., & Schenk-Hoppe, K. R. (2005). Evolutionary finance: introduction to the special issue. *Journal of Mathematical Economics*, 41(1-2), 1–5.
- Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review*, 34(1), 135–174. <http://doi.org/10.1111/dpr.12142>
- Hilbert, M. (2015). An Information Theoretic Decomposition of Fitness: Engineering the Communication Channels of Nature and Society (SSRN Scholarly Paper No. ID 2588146). Rochester, NY: *Social Science Research Network*. Retrieved from <http://papers.ssrn.com/abstract=2588146>
- Hirshleifer, J., & Riley, J. G. (1992). *The Analytics of Uncertainty and Information*. Cambridge University Press.
- Jensen, M. B., Johnson, B., Lorenz, E., & Lundvall, B. \AAke. (2007). Forms of knowledge and modes of innovation. *Research Policy*, 36(5), 680–693.
- Kelly, J. . (1956). A new interpretation of information rate. *Bell System Technical Journal*, 35(4), 917–926.
- Kolmogorov, A. (1941). Interpolation und Extrapolation von stationären zufälligen Folgen. *Bulletin Of the Academy of Sciences of the USSR, Series on Mathematics*(5), 3–14.
- Kolmogorov, A. N. (1968). Three approaches to the quantitative definition of information. *International Journal of Computer Mathematics*, 2(1-4), 157–168. <http://doi.org/10.1080/00207166808803030>
- Kramer, G. (1998). *Directed information for channels with feedback* (Doctoral Theses under J.L. Massey). Zurich: Swiss Federal Institute of Technology. Retrieved from <ftp://204.178.31.32/who/gkr/Papers/KramerThesis.ps.gz>
- Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. <http://doi.org/10.1214/aoms/1177729694>
- Kussell, E., & Leibler, S. (2005). Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. *Science*, 309(5743), 2075–2078. <http://doi.org/10.1126/science.1114383>
- Latané, H. A. (1959). Criteria for Choice Among Risky Ventures. *Journal of Political Economy*, 67(2), 144–155.
- Leung-Yan-Cheong, S., & Cover, T. (1978). Some equivalences between Shannon entropy and Kolmogorov complexity. *Information Theory, IEEE Transactions on*, 24(3), 331–338.
- Li, M., & Vitanyi, P. (1997). *An Introduction to Kolmogorov Complexity and Its Applications* (2nd ed.). New York: Springer.
- Lundvall, B., & Johnson, B. (1994). The Learning Economy. *Journal of Industry Studies*, 1(2). <http://doi.org/10.1080/13662719400000002>

- Maamar, H., Raj, A., & Dubnau, D. (2007). Noise in Gene Expression Determines Cell Fate in *Bacillus subtilis*. *Science*, 317(5837), 526–529. <http://doi.org/10.1126/science.1140818>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey & Company. Retrieved from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation
- Massey, J. L. (1990). Causality, feedback and directed information. In *Proceedings of the 1990 International Symposium on Information Theory and its Applications*. Waikiki, Hawai. Retrieved from http://www.isiweb.ee.ethz.ch/archive/massey_pub/pdf/BI532.pdf
- Massey, J. L., & Massey, P. C. (2005). Conservation of mutual and directed information. In *International Symposium on Information Theory, 2005. ISIT 2005. Proceedings* (pp. 157–158). <http://doi.org/10.1109/ISIT.2005.1523313>
- Maxwell, J. C. (1872). *Theory of heat*. Westport, Conn., Greenwood Press. Retrieved from <http://www.archive.org/details/theoryheat02maxwgoog>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work and think*. London: John Murray.
- Nelson, R. R., & Winter, S. G. (1985). *An Evolutionary Theory of Economic Change*. Belknap Press of Harvard University Press.
- Neumann, J. V., & Morgenstern, O. (1944). *Theory Of Games and Economic Behavior*. Princeton University Press. Retrieved from <http://archive.org/details/theoryofgamesand030098mbp>
- Nowak, M. A. (2006). *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press of Harvard University Press.
- OECD (Organisation for Economic Co-operation and Development). (2013). Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by “big data.” In *Supporting Investment in Knowledge Capital, Growth and Innovation* (pp. 319–356). Organisation for Economic Co-operation and Development. Retrieved from <http://www.oecd-ilibrary.org/content/chapter/9789264193307-12-en>
- Pabrai, M. (2007). *The Dhandho Investor: The Low - Risk Value Method to High Returns* (1st ed.). Wiley.
- Permuter, H. H., Kim, Y.-H., & Weissman, T. (2011). Interpretations of Directed Information in Portfolio Theory, Data Compression, and Hypothesis Testing. *IEEE Transactions on Information Theory*, 57(6), 3248–3259. <http://doi.org/10.1109/TIT.2011.2136270>
- Poincaré, H. (1908). Mathematical creation. *Reprint (2000) in Resonance*, 5(2), 85–94. <http://doi.org/10.1007/BF02838832>
- Rivoire, O., & Leibler, S. (2011). The Value of Information for Populations in Varying Environments. *Journal of Statistical Physics*, 142(6), 1124–1166. <http://doi.org/10.1007/s10955-011-0166-2>
- Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163–180. <http://doi.org/10.1177/0165551506070706>
- Saviotti, P. P. (1998). On the dynamics of appropriability, of tacit and of codified knowledge. *Research Policy*, 26(7-8), 843–856.
- Saviotti, P. P. (2004). Considerations about the Production and Utilization of Knowledge. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift Für Die Gesamte Staatswissenschaft*, 160(1), 100–121.
- Schumpeter, J. (1939). *Business Cycles: A Theoretical, Historical, And Statistical Analysis of the Capitalist Process*. New York: McGraw-Hill. Retrieved from http://classiques.uqac.ca/classiques/Schumpeter_joseph/business_cycles/schumpeter_business_cycles.pdf
- Shalizi, C. R., & Crutchfield, J. P. (2001). Computational Mechanics: Pattern and Prediction, Structure and Simplicity. *Journal of Statistical Physics*, 104(3-4), 817–879. <http://doi.org/10.1023/A:1010388907793>

- Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423, 623–656. <http://doi.org/10.1145/584091.584093>
- Solomonoff, R. J. (1964). A formal theory of inductive inference. Part I. *Information and Control*, 7(1), 1–22. [http://doi.org/10.1016/S0019-9958\(64\)90223-2](http://doi.org/10.1016/S0019-9958(64)90223-2)
- Spence, M. (1976). Informational Aspects of Market Structure: An Introduction. *The Quarterly Journal of Economics*, 90(4), 591–597. <http://doi.org/10.2307/1885323>
- Stearns, S. C. (2000). Daniel Bernoulli (1738): evolution and economics under risk. *Journal of Biosciences*, 25(3), 221–228.
- Szilard, L. (1929). über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Zeitschrift Für Physik A Hadrons and Nuclei*, 53(11), 840–856. <http://doi.org/10.1007/BF01341281>
- Tria, F., Loreto, V., Servedio, V. D. P., & Strogatz, S. H. (2014). The dynamics of correlated novelties. *Scientific Reports*, 4. <http://doi.org/10.1038/srep05890>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. <http://doi.org/10.1126/science.185.4157.1124>
- UNSD (United Nations Statistics Division). (2012). *UN COMTRADE database*. New York, NY: United Nations Statistics Division. Retrieved from <http://comtrade.un.org/>
- Weitzman, M. L. (1998). Recombinant Growth. *The Quarterly Journal of Economics*, 113(2), 331–360.
- Youn, H., Strumsky, D., Bettencourt, L. M. A., & Lobo, J. (2015). Invention as a combinatorial process: evidence from US patents. *Journal of The Royal Society Interface*, 12(106), 20150272. <http://doi.org/10.1098/rsif.2015.0272>
- Zeleny, M. (1986). Management Support Systems: Towards Integrated Knowledge Management. *Human Systems Management*, 7(1), 59–70.
- Zurek, W. H. (1989). Algorithmic randomness and physical entropy. *Physical Review A*, 40(8), 4731. <http://doi.org/10.1103/PhysRevA.40.4731>
- Zvonkin, A. K., & Levin, L. A. (1970). The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematics Surveys (Uspekhi Mat. Nauk)*, 25(6), 83–124.