STEM
ACADEMIC PRESS

INTERNATIONAL CONFERENCE ON COMPLEX SYSTEMS 2018

NEW ENGLAND COMPLEX SYSTEMS INSTITUTE

Alfredo J. Morales
Editor

IX International Conference on Complex Systems

Edited by Alfredo J. Morales
New England Complex Systems Institute
Cambridge, USA

July, 2018

Contents

A State-Space Modeling Framework for Engineering Blockchain-Enabled Economic Systems

Michael Zargham[1], Zixuan Zhang[2], Victor Preciado[2*†]

Abstract

Decentralized Ledger Technology (DLT), popularized by the Bitcoin network, aims to keep track of a ledger of valid transactions between agents of a virtual economy without the need of a central institution for coordination. In order to keep track of a faithful and accurate list of transactions, the ledger is broadcast and replicated across the machines in a peer-to-peer network. To enforce that the transactions in the ledger are valid (i.e., there is no negative balance or double spending), the network as a whole coordinates to accept or reject new transactions according to a set of rules aiming to detect and block the operation of malicious agents (i.e., Byzantine attacks). Consensus protocols are particularly important to coordinate the operation of the network, since they are used to reconcile potentially conflicting versions of the ledger. Regardless of the architecture and consensus mechanism used, the resulting economic networks remains largely similar, with economic agents driven by incentives under a set of rules. Due to the intense activity in this area, proper mathematical frameworks to model and analyze the behavior of blockchain-enabled systems are essential. In this paper, we address this need and provide the following contributions: (i) we establish a formal framework, using tools from dynamical systems theory, to mathematically describe the core concepts in blockchain-enabled networks, (ii) we apply this framework to the Bitcoin network and recover the key properties of the Bitcoin economic network, and (iii) we connect our modeling framework with powerful tools from control engineering, such as Lyapunov-like functions, to properly engineer economic systems with provable properties. Apart from the aforementioned contributions, the mathematical framework herein proposed lays a foundation for engineering more general economic systems build on emerging Turing complete networks, such as the Ethereum network, through which complex alternative economic models are being explored.

*1 Michael Zargham is Founder & CEO at BlockScience. Email: `zargham at block.science`

†2 Zixuan Zhang and Victor Preciado are with the Department of Electrical and Systems Engineering, University of Pennsylvania. Email: {`zixuanzh, preciado`} `at seas.upenn.edu`

1 Introduction

During the 2007-2008 global financial crisis, serious abuses by major financial institutions initiated a series of events resulting in a collapse of the loosely regulated financial network. This crash unveiled major weaknesses of traditional financial system and instigated a general feeling of distrust on banking institutions. In this context, decentralized economic systems based on cryptocurrencies, such as Bitcoin (*1*), were developed and launched by a group of cryptographic activists who believed in social change through censorship-resistant and privacy-enhancing technologies. Prior to Bitcoin, several attempts to establish digital currencies were made, including *b-money* by Dai (*2*), *hashcash* by Finney (*3*), and *bit gold* by Szabo (*4*). Instead of relying on large financial institutions for their operation, these cryptocurrencies proposed a decentralized economy in which a collection of economic agents coordinate through a peer-to-peer networks of computers via a blockchain protocol.

At a high level, blockchain aims to keep track of a ledger of valid transactions between agents of the economy without the need of a central institution for coordination. In order to keep track of a faithful and accurate list of transactions, the ledger is broadcast and replicated across all the machines in a peer-to-peer network. To enforce that the transactions in the ledger are valid (i.e., there is no negative balance or double spending), the network 'as a whole' coordinates to accept or reject new transactions according to a set of rules aiming to detect and block the operation of malicious agents (i.e., Byzantine attacks). Blockchain implements this idea by bunching together a group of new transactions that are added into a chain of blocks only if these transactions are validated by the peer-to-peer network. Consensus protocols are particularly important in this validation step, since they are commonly used to reconcile conflicting versions of the ledger. A particular protocol used to enforce consensus is *Proof of Work (PoW)* (*5*), currently used in Bitcoin and Ethereum. PoW is just one particular example of many other consensus protocols, such as the Practical Byzantine Fault Tolerant algorithm (PBFT) (*6*), Proof of Stake (PoS) (*7*), or Delegated Proof of Stake (DPoS) (*8*), to mention a few. These consensus protocols differ by their degree of decentralization, fault tolerance, throughput, and scalability (*9*).

It is worth mentioning that, beyond applications in cryptocurrency, blockchain can be used to store other forms of data on the network (*10*). This capability has inspired a plethora of protocols and applications aiming to, for example, certify the existence (*11*) and tracking of asset ownership (*12*). A notable application is Ethereum, a blockchain with a Turing-complete programming language (*13*). Ever since the inception of Ethereum, there has been an explosion of tokens sitting on top of the Ethereum network (*14–16*). Regardless of the architecture and consensus mechanism used, the resulting economic networks remains largely similar, with economic agents driven by incentives under a set of rules. Recent efforts have focused on implementing these currencies and other business logic such as decentralized exchanges through on-chain programs called smart contracts, first defined by Szabo in the 1990's (*17, 18*).

Due to the intense recent activity in this area, proper mathematical frameworks to model and analyze the behavior of blockchain-enabled systems are essential. In this direction, we find notable work by DeFigueiredo and Barr on path independence (*19*) and Dandekar et al. on credit networks (*20*). These works lay down the theoretical foundation for projects like Ripple (*21*) and Stellar (*22*). The main objective of this paper is to propose a new mathematical framework, based on tools from dynamical systems theory and control engineering, to model and analyze the function of blockchain-enabled systems. Before we provide more details about our particular modeling framework, let us describe the blockchain from a network point of view. In particular, the blockchain is comprised of two different networks: (a) a peer-to-peer computer network whose objective is to enforce and broadcast a valid state of the ledger, and (b) a network of valid transactions whose edges represent transactions and whose nodes are unique addresses used to encode the source and destination of a particular transaction. The main objective of the blockchain protocol is to keep a valid, up-to-date copy of the structure of network b) in all the nodes in network a).

As we mentioned above, we use tools from dynamical systems theory to model the function of the blockchain which is independent of the underlying computational infrastructure implemented. In particular, we borrow a modeling framework widely used in control engineering called *state space representation* (*23*). According to this framework, there is a set of abstract variables, called states, evolving over time (either continuous or discrete) according to a set of rules. In the discrete-time case, the evolution rules are described in terms of a first-order difference equation, in which the values of the states at a given time $t \in \mathbb{N}$ depend exclusively on the values of the states at time $t - 1$. Hence, given the initial values for the states at the origin of time (i.e., $t = 0$), it is possible to recover the states at any time $t > 0$ by solving this recursion. There is a rich mathematical theory to analyze state-space models, specially in the so-called linear case, in which the states at time t depend on the states at time $t-1$ according to a linear transformation. In this paper, we propose a linear state-space model of the blockchain network whose set of states represent transaction addresses. Notice that, as new transactions take place in the decentralized economy, the number of states in the model increases monotonically over time. This results in some technical difficulties that we overcome by proposing a linear time-expanding (LTE) state-space model which we will use to analyze the temporal behavior of the blockchain. As a reference case, we will model and analyze the evolution of the state in the public Bitcoin network using the aforementioned LTE modeling framework. Using tools from state-space theory (in particular, Lyapunov-like functions (*24*)), we will illustrate how to enforce a global property, namely, the total amount of currency in the system, using local state-transition rules.

The paper is structured as follows: After providing appropriate background, we describe the LTE state-space model in Section 2. In Section 3, we analyze the behavior of the Bitcoin network using our framework and provide tools for optimization and control enabled by our framework in Section4. We finalize with some conclusions and future considerations in Section 5.

2 Theoretical Framework

The following characterization attempts to create a useful abstraction over the properties of a blockchain network. It is not an attempt to describe how a blockchain network works, but to provide tools to engineer economic systems within such a network. All state variables are real-valued to make derivations more intuitive and straightforward.

2.1 Characterizing the Ledger

Definition 1 *The **Ledger State** is a shared data structure $\mathbf{B}(k) \in \mathcal{B}$ of a blockchain network which evolves in discrete time denoted by $k \in \mathbb{N}$. $\mathcal{B}$ denotes the domain space of all valid ledger states $\mathbf{B}(k)$ for any k.*

The Ledger State evolves in discrete time k, according to the rules and actions of all accounts. Each account is only accessible through the ownership of its private key. The Ledger State can thus be partitioned as $\mathbf{B}(k) = \bigtimes_{a \in \mathcal{A}_k} \mathbf{B}_a(k)$ where $\mathcal{A}_k$ denotes the set of all accounts at block k and $\bigtimes$ denotes the generalized cartesian product.

Definition 2 *An **Account** is a unique element in a ledger, identified by a public address $pk(a)$ with an associated private key sk. sk is required as a proof of right to both modify code defining the account and perform actions defined by the code.*

An account may contain code defining its state variables, external methods for interacting with other accounts, and other internally-used supporting methods. Since public keys are the unique identifiers of accounts, $PK = \mathcal{A}_k$ is the set of all public keys, similar to notation used in the cryptography community.

Definition 3 *A **Method** is a state transition function*

$$f : (\mathcal{U}, \mathcal{X}) \rightarrow \mathcal{X} \tag{1}$$

where $\mathcal{U} = \mathcal{U}(x)$ for $x \in \mathcal{X}$ is the set of legal actions.

Each method is defined by a particular account and is made available to a set of accounts. The set of methods provided by account i that are available to account j is denoted $\mathcal{F}_{(i,j)}$ with elements of the form

$$f_l : (\mathcal{U}_l, \mathcal{X}) \rightarrow \mathcal{X} \tag{2}$$

where l denotes the integer index of the functions within the set. The actions $\mathcal{U}_l(x)$ available to account j may depend explicitly on x_i.

A holder of the private key for account a can take any action by using any method in the set

$$\mathcal{F}_a = \bigcup_i \mathcal{F}_{(i,a)} \tag{3}$$

which has the associates space of actions

$$\mathcal{U}_a(x) = \bigcup_{l \in \mathcal{F}_a} \mathcal{U}_l(x) \tag{4}$$

for any $x \in \mathcal{X}$.

For the purpose of discussion, it is not required to further define internal operations that do not involve other accounts. Accounts are assumed to be passive explicitly and hence any internal state changes can be adequately accounted for in mappings $\mathcal{F}_a$.

The Ledger State directly related to account a at block k is

$$\mathbf{B}_a(k) = \{x_a(k), \mathcal{T}_a(k)\} \tag{5}$$

where $\mathcal{T}_a(k)$ is an ordered list of all transactions, involving account a in block k. Ordering is required as earlier changes to the state may influence the validity of later transactions in the sequence.

Definition 4 *A **Transaction** denoted by* $\mathtt{tx}$ *is the list of state changes caused by a discrete action* $u \in \mathcal{U}_a$ *for some account* $a \in A(k)$*. The initiating account is denoted* a_0 *and the other accounts whose states are impacted by the transaction are denoted by* a_i *for* $i \in 1, 2, 3, ...n$*; therefore the transaction itself is defined as the list* $\mathtt{tx} = [\Delta x_{a_0}, \Delta x_{a_1}, \ldots, \Delta x_{a_n}]$ *where* Δx_{a_i} *is the change in state of account* a_i *as a result of the action* u *represented by the transaction* $\mathtt{tx}$*.*

Under this notation, state transitions within a transaction are atomic and intermediate state transitions that may occur as part of a smart contract are not accounted for. A finer grained model would be required to handle that level of precision.

The simplest transaction occurs when $u \in \mathcal{U}_{a_0}$. The transaction only modifies the state of the initiating account by legally calling a method $f \in \mathcal{F}_{a_0}$ that does not change the state of any other accounts. The transaction is completely characterized by Δx_{a_0} and can be validated by evaluating $f(u, x)$ within the sequence of transactions $\mathcal{T}_{a_0}(k)$ and showing that $x_{a_0} + \Delta x_{a_0} = f(u, x)|_{a_0} \in \mathcal{X}_{a_0}$. The notation $f(\cdot)|_a$ denotes the element of the output of $f(\cdot)$ associated with account a.

A more interesting case is when account a_0 takes an action $u \in \mathcal{U}_{a_o}$ which changes the state of other accounts $a_i \in \mathcal{A}_k$. In order for $\mathtt{tx}$ to be a valid transaction the resulting state transitions for each other account must be valid. If the function f is implemented by calling other methods, the validity only holds if all methods are properly applied.

The rules on domain $\mathcal{X}$ are directly characterized by the range of state transition functions provided by accounts so it suffices to check the computation of $f(u, x)$ to assert the validity of each output account state:

$$\begin{aligned}
x_{a_0} + \Delta x_{a_0} = f_a(u, x)|_{a_0} &\in \mathcal{X}_{a_0}, \\
x_{a_1} + \Delta x_{a_1} = f_a(u, x)|_{a_1} &\in \mathcal{X}_{a_1}, \\
&\cdots \\
x_{a_n} + \Delta x_{a_n} = f_a(u, x)|_{a_n} &\in \mathcal{X}_{a_n}.
\end{aligned}$$

Definition 5 *A **Transaction Block** is an ordered list of transactions* $\mathbf{TX}(k) = [\mathbf{tx}_0, \mathbf{tx}_1, \ldots, \mathbf{tx}_m]$. *Since transactions are lists of account state changes, it can interpreted as a flat list. A block is valid if each individual transaction is computed correctly and the state change for each account after each sequential transaction is valid.*

Consider a block k with prior account states $x(k-1)$, the global account state update is given by

$$x(k) = x(k-1) + \Delta x(k) \tag{6}$$

where Δx is the global account state update achieved by all transactions in $\mathbf{TX}(k)$.

$$\Delta x_a(k) = x_a(k) - x_a(k-1) = \sum_i \Delta x_a^{(i)},\ \forall a \in \mathcal{A}_k \tag{7}$$

where the index i denotes the indices of all $\mathbf{tx} \in \mathcal{T}_a(k) \subseteq \mathbf{TX}(k)$, applying the definition $\mathcal{T}_a(k)$ is the set of transactions at block k including state changes to account a. The global state update

$$\Delta x(k) = x(k) - x(k-1) \tag{8}$$

can be completely characterized by $\mathbf{TX}(k)$ and evaluated account-wise according to equation (7).

Definition 6 *A **Block** is the ledger state* $\mathbf{B}(k)$ *at* k *which given the definitions above can be formally characterized as the pair*

$$\mathbf{B}(k) = (x(k), \mathbf{TX}(k)). \tag{9}$$

The underlying mechanisms that maintain the state of the blockchain do so by keeping the transaction history. This makes a direct connection to dynamical systems theory

$$x(K) = x(0) + \sum_{k=1}^{K} \Delta x(k) \tag{10}$$

where $x(0)$ denotes the initial state, also referred to as the genesis block.

Result 1 *The evolution of the Ledger State* $\mathbf{B}(k)$ *for* $k = 0, 1, 2, \ldots$ *is the trajectory of a discrete time second order networked system.*

It is a networked system comprised of accounts a with locally defined internal dynamics and rules for interacting according to the account Definition 2. It is a discrete second order system because the ledger state $\mathbf{B}(k)$ contains precisely the states $x_a(k)$ and the backward discrete derivatives $\Delta x_a(k) = x_a(k) - x_a(k-1)$ for all accounts $a \in \mathcal{A}_k$.

Any mechanism that can be implemented as an account under this framework provides an explicit contribution to the actions available to all other accounts within the system. The explicit characterization of an account and its subsequent state changes permits the estimation of changes in any utilities defined over the network state. Using the second order discrete networked system model, it is possible to both formally analyze the reachable state space and simulate the response to incentives with respect to a variety of behavioral assumptions.

2.2 Characterizing the Peer-to-Peer Network

Under this formal model there are two distinct concepts matching the term **Network**. The state space model defines the evolution of a network of interacting accounts. From this point of view, the economic network is a robotic network with agents represented by accounts, each of which has its own unique state space and action space defined in part by all of the other agents (accounts) in the network. The agent (account) states of all network participants and their backward discrete derivatives are visible to other agents and any external observer capable of querying the Ledger State.

Consideration of the external viewer brings attention to the other concept of a network which is required to model this system; the communication and computation network responsible for maintaining account states, computing state updates, verifying the validity of blocks of transactions, and to agree on the correct sequence of blocks when multiple valid sequences are available.

Definition 7 *A **Node** is a member of the Peer-to-Peer Network with the ability to broadcast a transaction* $\mathtt{tx}$ *for which it can prove control of the initiating account* a_0 *using the associated private key, and the ability to verify the validity of transactions broadcast by other nodes.*

Definition 8 *The **Peer-to-Peer Network** is the set of nodes* $j \in \mathcal{V}$*, participating in the communication and computation network, each maintaining a copy of the Ledger State* $\mathbf{B}_j(k)$ *and edges in this network represent communication between nodes.*

Note that each node j may have its belief of the Ledger State $\mathbf{B}_j(k)$ such that for any two nodes $\mathbf{B}_j(k) \neq \mathbf{B}_{j'}(k)$ for $j \neq j'$. However, It is guaranteed by the underlying cryptographic protocol that both $\mathbf{B}_j(k), \mathbf{B}_{j'}(k) \in \mathcal{B}$.

Definition 9 *A **Chain** is a valid sequence of Ledger States,* $\mathbf{C}(K) = [\mathbf{B}(k) \in \mathcal{B}$ *for* $k = 0, 1, \ldots, K] \in \mathcal{B}^{K+1}$ *where* $\mathbf{B}(0)$ *is the genesis block and* K *is the block height.*

The cryptographic protocol maintaining the ledger uses sequences of hashing functions to maintain a strict ordering on blocks such that any attempt to manipulate the history of the ledger state is immediately detectable by all nodes in the communication and computation network. Since the cryptographic protocol only accepts blocks for which all state transitions are defined by legal transactions, the chain is also guaranteed to contain a self consistent historical trajectory of the state space model, both states and derivatives, starting with the initial condition $x(0)$ as defined in the genesis block.

Returning to the issue of nodes maintaining valid but different chains $\mathbf{C}_j(K_j) \neq \mathbf{C}_{j'}(K_{j'})$ which conclude with Ledge States $\mathbf{B}_j(K_j), \mathbf{B}_{j'}(K_{j'}) \in \mathcal{B}$ where K_j and $K_{j'}$ may be different block heights.

Definition 10 *The **Consensus Protocol**,* $\mathbb{C}$ *is the process by which agents resolve inconsistency:*

$$\mathbb{C} : (\mathcal{C}, \mathcal{C}) \to \mathcal{C} \tag{11}$$

returning which of the two otherwise valid chains superseding the other.

The existence of such a function alone does not ensure the network does not fragment into partitions maintaining conflicting states. To further ensure consensus, the consensus protocol requires the following property.

Conjecture 1 *The Consensus protocol must impose a strict ordering on valid chains* $\mathbf{C} \in \mathcal{C}$*. It is sufficient that there exists a function*

$$\Psi : \mathcal{C} \to \mathbb{R} \tag{12}$$

such that for any $\mathbf{C}, \mathbf{C}' \in \mathcal{C}$

$$\mathbf{C} \neq \mathbf{C}' \implies \Psi(\mathbf{C}) \neq \Psi(\mathbf{C}'). \tag{13}$$

Two nodes may resolve their inconsistency by each setting their Chain to

$$\mathbf{C}^* = arg \max_{C \in \{\mathbf{C}, \mathbf{C}'\}} \Psi(C). \tag{14}$$

The formalism is consistent with the Nakomoto consensus paradigm where the function Ψ is the amount of work done to reach the current state in the competing chains. While technically it is possible for two chains to have exactly the same amount of work and still differ, but such a discrete event is a measure zero outcome in a continuous probability distribution, thus for the Bitcoin network using total work, equation (13) can be expected to hold with Probability 1.

For the purpose of the economic specification and subsequent design and analysis, it suffices to use any consensus protocol for which Conjecture 1 holds with Probability 1. Using proof schemes such as the one described above results in a lack of finality, meaning that there is always the possibility that another chain will supersede the one a node is maintaining. A consensus algorithm with **finality**, meaning that agreement on $\mathbf{B}(k)$ for block height K would not be reversible by some later observation, would be by definition Markovian. It would not be required to compare full trajectories $\mathbf{C}(K)$ to come to consensus over $\mathbf{B}(K)$. PoS-based consensus methods under development aim to achieve this property.

3 Bitcoin Reference Case

The public nature of the data in the Bitcoin economic network has made it a great candidate for research on financial flows. These models consider graphs of flows between accounts. This analysis will instead focus structurally on how the very simple rules about what constitutes a valid transaction result in well defined global properties.

3.1 Linear Time-Expanding Model

The Bitcoin economic network is defined over block height $k = 0, 1, 2, \ldots$, and there are $n_k = |\mathcal{A}_k|$ accounts at each block k with the additional caveat that $n_{k+1} \geqslant n_k$. For consistency of notation with dynamical models on networks, accounts will be referenced with indices $i \in \{1, \ldots, n_k\}$.

Definition 11 *A **Linear Time-Expanding (LTE)** system has a state space model in the form of a discrete time varying linear model with the dimension of the state space $x \in \mathbb{R}^{n_k}$ which is monotonically non-decreasing while the state update matrices vary only in n_k.*

Consider a canonical form discrete time linear time varying model:

$$x(k+1) = A_k x(k) + B_k u(k) \tag{15}$$

where $x(k) \in \mathbb{R}^{n_k}$. Under this framework $A_k \in \mathbb{R}^{n_{k+1} \times n_k}$, but since there are no internal dynamics

$$A_k = \begin{bmatrix} I_{n_k} \\ 0 \end{bmatrix} \tag{16}$$

where I_{n_k} is the identity matrix. The matrix B_k is an all-to-all incidence matrix encoding all possible sends $B_k \in \{0, 1, -1\}^{n_{k+1} \times m_k}$ where $m_k = n_{k+1} \cdot (n_k - 1) = |\mathcal{E}_k|$ and $u(k) \in \mathbb{R}^{m_k}$. The edge set is given by $\mathcal{E}_k = \mathcal{A}_k \times \mathcal{A}_{k+1}$ because flows must original in accounts that exist at time k.

$$[B_k]_{ie} = \begin{cases} 1 & \text{if } e = (j, i) \text{ for any } j \\ -1 & \text{if } e = (i, j) \text{ for any } j \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

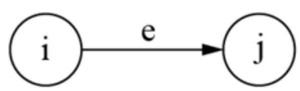


Figure 1: Illustration of incidence matrix. An edge e from i to j represents i sending a transaction to j. $B_{i,e} = 1$ and $B_{j,e} = -1$.

Result 2 *The system in equation* (15) *is an **Linear Time-Expanding** (LTE) system because for all k, A_k is an augmented identity matrix as defined in equation* (16) *and B_k is an all-to-all incidence matrix as defined in equation* (17).

The incidence matrix construction enforces the requirement of no double spends. Under this construction the local action $u_e(k) \in \mathcal{U}_e(k)$

$$\mathcal{U}_e(k) = \left\{ u_{e=(i,j)} \in \mathbb{R} | \sum_j u_{(i,j)} \leqslant x_i(k) \forall i \right\} \tag{18}$$

where equation (18) enforces the requirement that accounts cannot spend funds they do not have. Note that the requirement is locally enforceable, requiring only the balance of account i

and the other transactions to or from account i during block k. Viewed from the perspective of account i the local constraint a flow balance

$$x_i(k) + \sum_j u_{(j,i)}(k) - u_{(j,i)}(k) \geqslant 0. \tag{19}$$

In the practice, the transactions encoded by the inputs u are processed with a strict ordering that can be enforced with only the sender's state

$$u_{(i,j)} \leqslant x_i \tag{20}$$

as in definition 5. The model in equation (19) is a relaxation of the enforced requirement in equation (20); any block comprised of actions $u(k)$ that respect the individual transaction validity requirement equation (20) will satisfy the conservation law in equation (19). The relaxed equation in presented to demonstrate that the case of the Bitcoin network flow is in fact stronger than the conical network flow models in the controls literature.

3.2 Introducing Rewards

Since the genesis block contained an empty state these requirements would make for a trivial trajectory. In order to introduce funds into the economy, a driving function $M(k) = \mu_k v(k)$ is added:

$$x(k+1) = A_k x(k) + B_k u(k) + \mu_k v(k). \tag{21}$$

The function $M(k) \in \mathbb{R}^{n_{k+1}}$ is decomposed into a scheduled positive scalar reward $\mu_k \in \mathbb{R}_+$ and a stochastic vector $v(k) \in \mathbb{R}_+^{n_{k+1}}$ such that $\sum_i v_i(k) = 1$. The vector $v(k)$ denotes the distribution of the mining rewards across all accounts including potential allocation to new accounts or may be distributed by any arbitrary rule across an arbitrary subset of accounts, such as a mining pool.

Another key property of the Bitcoin is again recovered from out state space model. Define a scalar subspace of the state

$$y(k) = \mathbf{1}' x(k) = \sum_i x_i(k). \tag{22}$$

In order to under stand the behavior of this output function, consider

$$x(k) - A_{k-1} x(k-1) = B_k u(k) + \mu_k v(k) \tag{23}$$

and construct the state by summing the history of changes

$$\begin{aligned} x(K) =& A_{K-1} \cdots A_0 x(0) \\ &+ \sum_{k=1}^{K-1} A_{K-1} \cdots A_k \left(x(k) - A_{k-1} x(k-1) \right) \end{aligned} \tag{24}$$

becomes

$$\begin{aligned} x(K) =& A_{K-1}\cdots A_0 x(0) \\ &+ \sum_{k=1}^{K-1} A_{K-1}\cdots A_k\left(B_k u(k) + \mu_k v(k)\right). \end{aligned} \tag{25}$$

When this expression is used to compute

$$\begin{aligned} y(k) =& \mathbf{1}' x(K) \\ =& \mathbf{1}' A_{K-1}\cdots A_0 x(0) \\ &+ \sum_{k=1}^{K-1} \mathbf{1}' A_{K-1}\cdots A_k\left(B_k u(k) + \mu_k v(k)\right). \end{aligned} \tag{26}$$

Since $x(0) = 0$, it follows from equation (10) that

$$y(K) = \sum_{k=1} \mu_k \tag{27}$$

when one recalls that A_k is an augmented identity matrix, that $\mathbf{1}'v(k) = 1$ observes that B_k is an incidence matrix: $\mathbf{1}'Bu = 0$ for all u.

In the case of Bitcoin the mining rewards are on a convergent schedule ensuring the maximum total supply

$$y_\infty = \lim_{k\to\infty} y(k) = \sum_{k=1}^{\infty} \mu_k \tag{28}$$

converges to the desired quantity.

In the Bitcoin network the mining rewards are defined over $i = (1, \ldots, 32)$ halving intervals $r_i = (k_0, k_1, \ldots, k_{209999})$ each including 210000 blocks resulting in

$$\mu_k = \frac{\left\lfloor \frac{50\cdot 10^8}{2^i} \right\rfloor}{10^8} \text{where } k \in r_i. \tag{29}$$

After the 32^{nd} interval the minted block rewards cease and the total quantity of Bitcoin is conserved. By computing the sum over the intervals, the final sum of Bitcoin $y_\infty = 20999999.9769$, generally quoted as 21 million BTC. This does not account for the potential loss of control of accounts with Bitcoin balances which reduces the effective supply.

3.3 Globally Invariant Properties from Local Rules

While the Bitcoin economic network is a somewhat trivial system to study from a dynamical systems perspective, it is actually much like biological coordination models, an example of complex global behaviors emerging from simple local rules. The critical elements of the above example are:

- The trajectory of the system is defined entirely in terms of its state transitions and the initial conditions.
- The dynamical system model remains structurally invariant even as the number of account grows unbounded.
- The inputs or actions are completely under the control of local agents called accounts.
- The set of legal actions for each agent are defined and verifiable with information local to those agents states.
- The definitions of the local legal actions provide properties that are consistent with suitability as a financial ledger of record.
- The driving function combined with the legal actions guarantee that a low dimensional global property is enforced throughout the entire trajectory.

The most powerful part about this characterization is that the system literally tracks a desired property $y(k) = \sum_k \mu_k$ for the entire trajectory, in fact in any valid trajectory, with no assumptions about the actions of the individual agents. This indicates that it is proper to think of blockchain-enabled economic systems as engineered economies where it is possible to encode the legal state transitions in such a manner as to mathematically ensure the emergence of a low dimensional global properties.

4 Value functions, Invariants, Optimization and Control

Where Bitcoin was precisely characterized by the set of hard coded rules, more general Turing complete networks such as the Ethereum network are capable of much richer state spaces and legal state transitions. In this section, the concepts explored in the Bitcoin section will be generalized for use in characterizing a more general system state.

Consider that each account may contribute a set of state variables to each other account; the cardinality of the state space grows super-linearly. At first pass this may seem to make the system difficult to work with, the opposite may in fact be true. The creator of an account is likely to have goals specific to the states it instantiates and can define precisely the functions which mutate those states including relations to other states. In particular, the goals for the system are likely to exist in a much lower dimension than the state itself and thus it is totally reasonable to provide an extremely high degree of freedom to the agents in the system while formally enforcing those goals.

For the purpose of this characterization, consider a smart contract account α in Turing complete computational network.

Definition 12 *The account α defines a state variable z_a for all accounts $a \in \mathcal{A}_k$, for which the global state*

$$z(k) = \{z_a \in \mathcal{Z}_a | \forall a \in \mathcal{A}_k\} \in \mathcal{Z} \subset \mathcal{X},$$

where $\mathcal{Z}_a$ is an arbitrary state space contributed to every account a by account α.

Be aware that $z(k)$ includes part of the state of every agent $x_a(k)$, not the local states of agent α which is denoted $x_\alpha(k)$. Furthermore, the account α provides a set of methods that allow any account to modify the global state $z(k)$.

Definition 13 *The set of functions provided by account α is*

$$\mathcal{F}_\alpha = \{f_l : (\mathcal{U}_l, \mathcal{Z}) \to \mathcal{Z}) | \forall l\} \tag{30}$$

where $\mathcal{U}_l$ is the set of actions or inputs to each function f_l and $\mathcal{Z} \subset \mathcal{X}$ is the domain.

These functions define the feasible trajectories in the state of the system. Engineering these systems involves formally determining these functions in order to ensure system level properties are met. System level properties can be characterized by scalar functions of the state. The system becomes increasingly more complex as contracts are introduced to build on each others state variables.

However, building on previous contracts is accomplished by calling methods previously available made available to mutate their states, it suffices for now to examine properties in the context of a single contract; that is to say, assume all states and state transitions are provided by account one account α. Then $z(k) = x(k)$ be the full state with domain $\mathcal{Z} = \mathcal{X}$ and $\mathcal{F}_\alpha = \mathcal{F}$ accounts for all valid state transitions.

4.1 Value Functions

Value functions are output functions engineered to encode desired global properties. By limiting them to be positive scalars additional mathematical equipment can be brought to bear regarding convergence properties.

Definition 14 *A Value function is a non-negative scalar function of the state of the economic network*

$$V : \mathcal{X} \to \mathbb{R}_+$$

encoding some property.

These functions are measures of properties and need to be constructed for a particular property. The first case, that will be considered is when $V(x) = c$ encodes an equality constraint. It is possible to achieve such a constraint within the network by proving an invariant.

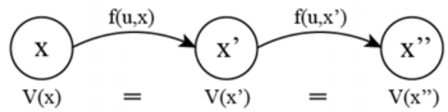


Figure 2: Illustration of invariant properties. V is invariant under all $f(u, x)$ for all valid actions u.

Definition 15 *An invariant property $V(x) = c$ is ensured by verifying that*

$$V(f_l(u, x)) = V(x)$$

for all functions $f_l \in \mathcal{F}$, actions $u \in \mathcal{U}_l$ and all legal states $x \in \mathcal{X}$.

In the definition above, all states are considered but as needed the same definition can be used over an arbitrary subset set of the state space. In practice, the engineer will find it prudent to work with a minimal subset of the state.

The fewer states which actually appear in the domain of $V(x)$ and $f_l(x)$ the more easily the invariant property in Definition 15 will be to demonstrate. Further note that the construction of $V(x)$ allows a large class of set based constraints.

In the Bitcoin example the positivity requirement could be encoded as $V(x) = -\sum_i \min\{0, x_i\}$=0. This function is well defined for all $x \in \mathbb{R}^n$, but is equal to zero only if every balance is positive. The fact that all send actions u_e require that $x_i > u_e$ where $e = (i, j)$ guarantees this invariant.

4.2 Maximization

A similar construction will allow other dynamics to be enforced. In the case of a value function $V(x)$ which the economic network wishes to collectively maximize, the following construction is sufficient.

Definition 16 *A global maximization property over value function $V(x)$ is guaranteed by the requirement*

$$V(f_l(u, x)) \geqslant V(x)$$

for all functions $f_l \in \mathcal{F}$, actions $u \in \mathcal{U}_l$ and all legal states $x \in \mathcal{X}$.

Observe that this does not guarantee infinite growth of $V(x)$ but rather that $V(x)$ cannot be reduced by any legal action provided. As in the case of invariant properties, it is prudent to restrict the subset of the state which your smart contract maintains as such arguments may break down in the case where other accounts provide functions that operate on these state variables in a manner which violates the property. A strict version of this mechanism can be imposed if

$$V(f_l(u, x)) \geqslant (1 + \epsilon)V(x) \quad (31)$$

for $\epsilon > 0$. This is the type of construct that can be used to ensure that perpetual revenues can be produced by the ongoing use of a network. However, if these profits are not fairly aligned with the services provided in facilitating transactions, transactions may halt and from the perspective of this system, time will have stopped and there will be no more revenue.

4.3 Value Functions as Lyapunov Functions

A Lyapunov function is analogous to a potential function and is a tool for proving stability around an equilibrium in a dynamical system. In the case of blockchain-enabled economic systems, the dynamics are entirely defined by the states and methods provided by accounts. As a result one need not discover Lyapunov functions to prove convergence to desired properties but can instead engineer the functions provided working backwards from a value function of choice. This approach is more powerful even than invariant simple invariant properties because the system will persistently move towards the desired state even if some unforeseen action causes a change in $V(x)$.

Definition 17 *An exponentially globally convergent property $V(x) = 0$ is guaranteed by the requirement*

$$V(f_l(u, x)) \leqslant \gamma V(x)$$

for all functions $f_l \in \mathcal{F}$, actions $u \in \mathcal{U}_l$ and all legal states $x \in \mathcal{X}$ where $\gamma \in [0, 1)$ is the exponential convergence rate.

These results are simple consequences of the Brouwer fixed-point theorem (*25*) provided that the space $\mathcal{X}$ is a convex compact set. Note that creating such a construction is non-trivial, but it formally ensures that all legal actions will drive the system towards the target subset of the state $\{x \in \mathcal{X} | V(x) = 0\}$. This construction further implies that crypto-economic networks are well suited to Lyapunov style control.

Definition 18 *Consider an output function $y = g(x)$ and target trajectory $y(k)$, define a Value function $V(x)$ such that $V(x) = 0$ if and only if $g(x) = y(k)$. This construction will be denoted a Lyapunov controller provided that*

$$V(f_l(u, x)) \leqslant \gamma V(x)$$

for all functions $f_l \in \mathcal{F}$, actions $u \in \mathcal{U}_l$ and all legal states $x \in \mathcal{X}$ where $\gamma \in [0, 1)$ is the exponential convergence rate.

The resulting system will under all legal actions continuously attempt to converge to the subset of the state space here $g(x) = y(k)$; note that the exponential convergent rate must be faster than the evolution of the target signal $y(k)$ in order for this process to track the desired target within a reasonable neighborhood.

Returning to the Bitcoin example, enforcing the invariant $y(K) = \sum_i x_i(K) = \sum_{k=1}^{K} \mu_k$ is actually an explicit case of controlling the system to track a low dimensional target state that evolves over time using a positive scalar value function. No convergence argument as in Definition 18 because the value function was invariant under all legal state transitions as in Definition 15

It will be interesting to see what types of emergent coordination could be engineered in economic systems by leveraging Lyapunov functions in the design of the legal action sets. It is expected that more general conservation equations may encoded to properly account for financial value changing forms, similar to how conservation of energy is handled in physical systems.

5 Conclusion and Future Considerations

This document builds a bridge between dynamical systems theory and blockchain-enabled economic systems. In particular, we propose a state-space representation of the economic system in terms of linear time-expanding system. This novel representation allows us to use a plethora of powerful tools developed in the context of control theory for the analysis and design of blockchain-enabled systems. In particular, we propose the use of Lyapunov-like functions, originally developed in the context of dynamical systems, to properly engineer economic systems with provable properties. A few comments are in order: First, no direct attention is paid to the actions of individuals within the network and all guarantees proposed are derived from the spaces of actions those individuals might take. Second, it may not always be possible to sufficiently limit the action spaces to ensure desired properties without considering individual incentives. In addition to the definitions provided above the author posits that value functions can be used to analyze the alignment of incentives of individuals actions with global objectives by comparing the gradients of local incentive functions with those of the global value functions.

A key aspect of our future research will include the design and development of simple on-chain controllers and the empirical study of the evolution of the states in the Ethereum ecosystem for comparison with theoretical principles outlined above. Tools for streamlined access to data from blockchain economic networks are under development.

References

1. Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. `https://bitcoin.org/bitcoin.pdf`, Dec 2008. Accessed: 2015-07-01.

2. Wei Dai. bmoney. `http://www.weidai.com/bmoney.txt`, 1998. Accessed: 2016-04-31.

3. Adam Back et al. Hashcash-a denial of service counter-measure. `http://www.hashcash.org/papers/hashcash.pdf`, 2002. Accessed: 2016-03-09.

4. Nick Szabo. Bit gold. http://unenumerated.blogspot.co.at/2005/12/bit-gold.html, 2005. Accessed: 2016-04-31.

5. Iddo Bentov, Ariel Gabizon, and Alex Mizrahi. Cryptocurrencies without proof of work. http://arxiv.org/pdf/1406.5694.pdf, 2014. Accessed: 2016-03-09.

6. Miguel Castro, Barbara Liskov, et al. Practical byzantine fault tolerance. In *OSDI*, volume 99, pages 173–186, 1999.

7. Ethereum community. Proof of stake: Faq. https://github.com/ethereum/wiki/wiki/Proof-of-Stake-FAQ.

8. Bitshares community. Delegated proof-of-stake consensus. https://bitshares.org/technology/delegated-proof-of-stake-consensus/.

9. Zhang Zhe Wang Xiangwei Chen Qijun Du Mingxiao, Ma Xiaofeng. A review on consensus algorithm of blockchain. *IEEE International Conference on Systems, Man, and Cybernetics*, 2017.

10. Massimo Bartoletti and Livio Pompianu. An analysis of bitcoin op_return metadata. https://arxiv.org/pdf/1702.01024.pdf, 2017. Accessed: 2017-03-09.

11. Factom. A practical blockchain solution for preserving, ensuring and validating digital assets. https://www.factom.com/.

12. Omni Layer. Omni layer, an open-sourced, fully-decentralized asset platform on the bitcoin blockchain. http://www.omnilayer.org/.

13. Vitalik Buterin. Ethereum: A next-generation smart contract and decentralized application platform. https://github.com/ethereum/wiki/wiki/White-Paper, 2014. Accessed: 2016-08-22.

14. Brave Software. Basic attention token (bat) blockchain based digital advertising. https://www.basicattentiontoken.org/BasicAttentionTokenWhitePaper-4.pdf, 3 2018.

15. Will Warren, Amir Bandeali. 0x: An open protocol for decentralized exchange on the ethereum blockchain. https://icowhitepapers.info/data/0x_white_paper.pdf.

16. Grid+. Grid+: welcome to the future of energy. https://gridplus.io/assets/Gridwhitepaper.pdf.

17. Nick Szabo. Smart contracts. http://szabo.best.vwh.net/smart.contracts.html, 1994. Accessed: 2016-08-22.

18. Szabo Nick. Smart contracts: Building blocks for digital markets. http://szabo.best.vwh.net/smart_contracts_2.html, 1996. (Accessed on 08/22/2016).

19. D. B. DeFigueiredo and E. T. Barr. Trustdavis: a non-exploitable online reputation system. In *Seventh IEEE International Conference on E-Commerce Technology (CEC'05)*, pages 274–283, July 2005.

20. Pranav Dandekar, Ashish Goel, Ramesh Govindan, and Ian Post. Liquidity in credit networks: A little trust goes a long way. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, EC '11, pages 147–156, New York, NY, USA, 2011. ACM.

21. David Schwartz, Noah Youngs, and Arthur Britto. The ripple protocol consensus algorithm. https://ripple.com/files/ripple_consensus_whitepaper.pdf, 2014. Accessed: 2016-08-08.

22. David Mazieres. The stellar consensus protocol: A federated model for internet-level consensus. https://www.stellar.org/papers/stellar-consensus-protocol.pdf, 2015. Accessed: 2016-08-01.

23. Eduardo D Sontag. *Mathematical control theory: deterministic finite dimensional systems*, volume 6. Springer Science & Business Media, 2013.

24. Hassan K Khalil. Nonlinear systems. *Prentice-Hall, New Jersey*, 2(5):5–1, 1996.

25. David Roger Smart. *Fixed point theorems*, volume 66. CUP Archive, 1980.

The Emergence of Trust and Value in Public Blockchain Networks

Michael D. Norman*, Yiannis G. Karavas, Harvey Reed

The MITRE Corporation, Bedford, MA 01730
*E-mail: mnorman@mitre.org

Public blockchain networks, such as Bitcoin and Ethereum, appear to be complex systems, due to the emergence of certain properties which only become observable at a system level from the choices made by a set of decentralized participants. Two of the most salient properties of these complex public blockchain networks are trust and value. We propose that a complex systems-based perspective should be applied to public blockchain networks and offer a number of possible applications of complexity science in this space.

1 Introduction

In this paper, we will present logical evidence that the properties of trust and value that are considered to be fundamental to public blockchain (PB) design are in fact emergent in nature, and thus do not exist in any single system component.

The existence of trust is demonstrated by the fact that, despite the natural human psychological desire to store wealth in tangible assets such as gold, Bitcoin, a completely intangible asset, is now used by millions to store vast amounts of wealth. The existence of value is demonstrated in the monetary value that Bitcoin tokens have in an open market.

The underlying technologies and principles composing the infrastructure of blockchain technology, such as cryptography, decentralization, and linked lists, are not novel in and of themselves. The unique combination of these technologies, culminating in the creation of blockchain implementations, leads to the formation of public, immutable, distributed ledgers such as Bitcoin. Blockchain networks such as Bitcoin appear to be complex systems due to the emergence of certain properties which only become observable at a system level: two of which are trust and value.

PB networks, no matter what the details of their implementations, are composed of decentralized sets of nodes. Decentralization is considered by the PB community to be a source of system-wide resilience, due in part to the network's ability to adapt to a dynamically changing population. What often fails to get attention is the impact of human decision making on a PB's viability, and how trust and value emerge from a collective decision-making dynamic, rather than from traditional top-down sources. The fact that each one of the aforementioned nodes is (likely) operated by a separate individual capable of making a decision to participate (or not) in the system's infrastructure at any given moment, coupled with the fact the system's performance and value are causally linked to the number of participating nodes, justifies treating a PB as a complex adaptive system.

Consider, as a mental exercise, an Ethereum network with only one participating entity. Does trust exist? What about when there are three equally influential entities? How many equally-influential entities do we have to have participating before that trust actually has value? These types of questions, ones related to nonlinearity and tipping points, are the ones that are typically the drivers of modeling emergent phenomena (*1*).

If we consider a PB to be a complex adaptive system, then we can apply the tools of complexity science, sometimes referred to as Complex Systems Engineering (*2*) to improve PB designs. In what follows, we will discuss the details of this perspective.

2 Definitions

Emergence: The notion of emergence has existed since the time of Ancient Greece. The modern term for it, emergence, was first used and subsequently defined by the 19th century English philosopher and critic G.H. Lewes who wrote:

> ”Every resultant is either a sum or a difference of the co-operant forces; their sum, when their directions are the same their difference, when their directions are contrary. Further, every resultant is clearly traceable in its components, because these are homogeneous and commensurable. It is otherwise with emergents, when, instead of adding measurable motion to measurable motion, or things of one kind to other individuals of their kind, there is a co-operation of things of unlike kinds. The emergent is unlike its components insofar as these are incommensurable, and it cannot be reduced to their sum or their difference” (*3*).

In short, this definition suggests that for a complex system, the whole is greater than the sum of its parts due to the manifestation of properties at a system level. An emergent property is not visible upon inspecting an individual component of the system. It only becomes apparent when the constituent components of the system coalesce and interact dynamically.

Decentralization: The transfer and diffusion of power and/or authority from a centralized, often singular, entity to several independent entities.

Human Incentives: Traditionally defined, an incentive is something that motivates one to perform a certain action. Specifically, in the context of blockchain technology, a human incentive is a reward, usually with an associated monetary value, that is distributed to human operators of nodes in the blockchain as a form of reimbursement for expending capital in the form of computers and electricity while securing the blockchain (i.e., mining or staking).

Applied Complexity Science:. The act of engineering complex adaptive systems using the methods of complexity science (*1*). The goal of these systems is to produce non-pathological emergent behavior. This perspective stands in stark contrast to one employed by traditional systems engineering, which seeks to remove complexity from the system and avoid emergent behavior. Practitioners of Applied Complexity Science seek to leverage the emergent properties of complex systems.

3 Emergence of Trust

Traditional cyber-attack vectors tend to be weighted towards the strategy of vulnerability exploitation (*4*). PB technology is designed to shift these traditional attack vectors to be necessarily and disproportionately weighted, if not almost exclusively weighted, towards the strategy of brute forcing. PB technology accomplishes this shifting of attack vectors through its archetypal decentralized design which forces potential attackers to successfully execute what is known as a 51% attack (*5*). Carrying out an attack of this type is contingent upon the attackers ability to successfully compromise over half the nodes in the blockchain. Control of a majority of the nodes in a PB would allow one to effect any desired change on the blockchain. By necessitating that potential perpetrators marshal vast resources prior to contemplating such an attack, this technology raises the cost of cyber-war for potential aggressors in a manner that creates immense power asymmetry and ultimately disincentivizes many potential attacks. In being designed to engender such substantial disincentives to attack it, while on the other hand offering incentives to participate in and support it, PB technology effectively generates a field of trust and security which grows more fortified as the number of participants in the system increases.

In observing a single node of a PB, one of the most integral properties of this technology remains elusive if not impossible to identify. Only upon more thorough examination and/or consideration of a larger group of nodes, does the significant property mentioned above, of mutual trust between nodes, begin to manifest itself. This emergent feature, observed in implemented PB networks, has become a virtually inseparable and vital part of the definition of blockchain technology as it is currently understood. PB technology, at its core, is a cleverly designed assemblage of pre-existing technologies engineered to channel human nature, in a manner such that the composite effect of many selfish nodes competing for the same reward, creates an expansive, systemic consensus spanning the nodes of the given network. Progress is made in reaching this state of consensus as these nodes, fueled by monetary incentive, pursue their selfish goal until one node successfully finds, or rather mines, a new block in the blockchain.

It is important to note that a block can only be verified if the node performing the computations uses an accepted set of rules for the verification process; rules which are themselves determined by consensus on the blockchain (*6*). These rules, such as hashing algorithms, maximum transaction count per block and many more are designed and/or chosen not only for standardization purposes, but also, more importantly, for ensuring that malicious nodes attempting to add false or selfish transactions within a block will fail verification. Utilizing this schema, blockchains render attempts at dishonest gains motivated by selfishness ineffective; thus, malicious nodes are denied mining rewards while continuing to waste valuable resources in mining computations, resulting in a net loss of value. Subsequently, the only feasible possibility for a would-be malicious node, or any node in general, to gain profit, is by channeling this selfish drive towards mining the next block as quickly as possible, in an honest fashion. Once a block is mined, it must be propagated throughout the network and verified by more than 50% of the nodes comprising the blockchain (*5*). If this majority verification stage succeeds, the blockchain is said to be in consensus. For the blockchain to function properly, this state of consensus is assumed to be the truth, meaning that no malicious nodes succeeded in propagating unauthorized transactions in the most recent block. This assumption of truth is possible, if and only if, any given node, n, in a blockchain of N nodes, trusts that at least [floor(N/2) + 1] nodes that form the blockchain it is a part of, are operating according to the agreed upon rules used for block verification. Given that the design of a typical PB shifts traditional cyber-attack vectors to make a 51% attack extremely unfeasible, this assumption of consensus and truth being equivalent is a safe assumption in the context of a blockchain. In this way, trust emerges, rather unexpectedly and paradoxically, yet necessarily, from a profusion of selfish entities. Since this trust is only observable in a system of nodes, one node alone being inadequate, it follows that this trust is in fact an emergent property of a complex PB system. Ironically enough, trust emerges over trust-less, decentralized, and privately-owned infrastructure.

In a complex system, competition at one scale enables cooperation at another (*7*). In PBs, we see this principle of human dynamics augmented technologically: in proof-of-work, as discussed, miners compete to form valid blocks, and this competition enables the cooperative behavior required to conduct financial transactions.

By creating a competition that appeals to selfish human nature through incentive, decentralization and trust are preserved. Trust is preserved because of an immutable ledger not controlled by any single entity. The competitors compete because this trust has value. This is interesting because the design of the system requires one to cross the system's boundaries: the blockchain is not able to know anything outside of itself, so its initial designers had no idea if the token generation mechanism would actually take. The prerequisite marketing efforts and network effects required to kick-start a PB are nontrivial to say the least, and should be considered a part of PB design. Once large groups of networked human decision-makers are included in the design of the system, the tools of complexity science can be readily applied. For example, complexity scientists could create agent-based models of social network effects to explore the qualities of a given PB design.

4 Emergence of Value

In its most basic form, economic value is derived from the perceived benefit that a good or service provides (*8*). The perceived benefit comes from the favorable properties of a PB, including its robustness to both targeted and random attacks which is derived from its decentralized nature. These beneficial properties, in turn, foster a favorable perception which leads to the generation of demand for the asset in question. Finally, this increase in demand inflates the asset's value. Once it is established that there exists a demand for a product, industry, in whatever form necessary, arises to generate a supply. Thereafter, the assets market value is determined, in large part, by the ratio of supply to demand. For the purposes of clarity and conciseness, this section will focus on the Bitcoin blockchain as an example for demonstrating the proposed possibility of the emergence of value in an implemented PB network. Bitcoins maximum supply is decidedly limited to 21 million tokens and can therefore be considered a constant. This leaves only one variable in the context of its economic valuation, demand. Given these unique circumstances, the terms value and demand can be used interchangeably when considering Bitcoin. There are many factors that can be considered to contribute to the demand for an asset such as Bitcoin. These factors include, but are not limited to, its security/emergent trust, liquidity and decentralization.

Cybersecurity has indisputably become a subject of paramount concern when designing software applications in many sectors of society, including banking, e-commerce, and national security. Furthermore, the supply of skilled workers in this field is failing to meet the demand, thus increasing the value of the cybersecurity measures provided by these workers (*9*) (*10*). For these reasons, security, such as that provided by Bitcoin as a consequence of its emergent property of trust, can indeed be considered valuable. Additionally, the value of Bitcoins security is compounded by the fact that it has proven unassailable and completely resilient to any form of attack to date. While Bitcoin's impregnability is not an absolute certainty as time goes to infinity, it is an empirical property which has persisted and continues to persist since its inception. Such a resilient and robust application of a defensive cybersecurity measure has rarely, if ever, been observed for such an extended period of time in any form since the conception of cyberwarfare. The token on the Bitcoin blockchain derives its value, in part, from this extremely robust security which is responsible for preserving the validity of the state of the blockchain and the transactions transpiring within it.

Although security against malicious activity is very valuable, especially in an increasingly digitized society, a PBs security alone is not sufficient to lend value to assets such as Bitcoin. A transaction in which a node receives Bitcoin from another is the changing of a single variable in software at minimum. This alteration of a variable in software can be described physically as the flipping of one or more computer bits in a machine's memory. Therefore, if security alone is enough to create the value associated with assets on a blockchain, the argument can be posed that a document with a large, random number written to it (flipping bits), saving the file in such a way that it can no longer be altered, encrypting it on an external drive and locking that drive in a physical safe should be very valuable due to its extreme security; this, of course, is not necessarily the case. In the act of securing this supposed asset, a trade-off must be made between its property of increased security and its properties of accessibility, liquidity and third party visibility. For that asset to be used, it must first be stripped of its security, consequently leaving it vulnerable to attack, transferred to the intended recipient and then re-secured. Bitcoin is unique in that transactions committed on its blockchain do not require this trade-off in order to execute. The token asset remains secure whether it is in an account or being transferred between accounts on the blockchain. Furthermore, because a blockchain such as Bitcoin can be, and is a distributed public ledger (*11*) - due to its relative immutability - all transactions and states of token assets are visible to all parties. This maintained state of security during storage and even during occurrences of transaction, allows for constant accessibility to the token assets since they are not required to be stored in a manner which temporally or spatially restricts an account owner's access to them. This accessibility leads to increased liquidity in a network of nodes. Accessibility itself is not an emergent property since it can be observed even in a single isolated node. However, ready accessibility does allow for the emergence of liquidity in a network. Liquidity, or the measure of difficulty of an entity's ability to exchange a given asset for another in the economic ecosystem in question (*12*), requires the existence of more than one entity since the act of exchange implies the existence of more than one entity. Therefore, if only one node comprised the Bitcoin network, accessibility becomes irrelevant since no other nodes would exist with which to transact and thus the measure of liquidity becomes meaningless. Hence, a blockchain network, allows not only for the property of security to emerge, but also for the property of liquidity to emerge. This unique, rare and economically attractive combination of properties - sustained, nearly impregnable security, liquidity, and visibility - collectively create demand for Bitcoin, thus increasing its value. This value seems to be an emergent property since it is derived from properties which are themselves emergent.

The decentralized nature of PB technology also contributes to the value of assets hosted on the blockchain. A common motif in the progress of human civilization has been the empowerment of the individual. This motif can be observed in its different forms throughout history, for example, in periods such as the Civil Rights movement during the 20th century and the Age of Enlightenment in 18th century Europe. One of the driving forces behind the Enlightenment is similar to that of blockchain technology decentralization (*13*). Attempts at decentralization during the Enlightenment took the form of a movement away from centralized forms of government, such as monarchies, to more decentralized and representative forms of government, such as republics. These representative governments naturally led to increased power and rights for the individual. This historical example demonstrates the ability of decentralization to satisfy an innate human desire for individual empowerment and control. Blockchain technology offers this same empowerment to the individual through its decentralized nature by allowing one to maintain full control over ones assets as opposed to subscribing to the traditional model of protecting assets in centralized institutions such as a banks, although this option is not precluded at any point; an individual could always pay a service to hold their private key(s) if they were not comfortable or technically competent enough to do so. As this technology matures to allow the creation of decentralized applications (dapps) (*14*) to be hosted on PBs, the definition of 'asset' in the context of a PB broadens as well. Just as Bitcoin allows the user to maintain full control over their monetary assets in the form of Bitcoin, dapps offer a user the ability to maintain full and secure control over their digital information used within the dapp. For example, if a decentralized version of Facebook was to be hosted on a blockchain as a dapp, users would have full control over their personal information used within the application. Models such as this not only neutralize the threat of personal information being stolen off centralized servers, but also alleviate paranoia and concerns that personal data is being unlawfully abused and/or sold by the corporations that maintain and own the servers hosting centralized applications. These two major advantages of the decentralized nature of PB contribute to the empowerment of the individual and give the individual proper control over their personal property, hence adding value to blockchain technology as well as to the assets housed on PBs themselves.

5 Valuation Methods

As in the previous section, for the purposes of clarity and conciseness, this section will focus on the Bitcoin blockchain as an example for considering various blockchain valuation concepts.

The valuation of Bitcoin has been an extremely volatile affair over the past few years. Currently, many economists and experts indicate that defining an underlying intrinsic value for Bitcoin has proven elusive (*15*). This elusiveness is to be expected, however, since Bitcoin and the underlying technology, blockchain, is a relatively new phenomenon, the development of which has been tremendously dynamic and fast paced. In order to arrive at a solid understanding of Bitcoins value, it is necessary to understand the underlying technology upon which Bitcoin is built. Due to the nature of PB technologys dynamic and ever-changing state, determining the true value of Bitcoin seems impossible for the time being. A fitting parallel may be seen in the commercialization and early mass adoption of the internet. During this period in the mid to late 1990s, there was no solid consensus in sentiment regarding the utility and value the internet could provide. This confusion in sentiment was likely, in part, due to a lack of understanding of what the internet was capable of and what utility it could provide. As the technology matured and understanding broadened, sentiment began to turn positive as the internets potential, and in turn intrinsic value, became more apparent. Today the internet has indisputably added immeasurable value to society. Likewise, the development stages of blockchain technology and Bitcoin are in their infancy which is why determining their intrinsic value has proven difficult. Considering this unsettledness in the development of blockchain technology, other methods of valuation must be employed. One possible alternative method is predicated upon the premise that Bitcoins value stems from its security.

Spending on cybersecurity in the United States has risen from $27.4 Billion in 2010 to $66 Billion in 2018 (*16*). The majority of this cost comes from the development of security measures against cyber attacks. Other factors contributing to this cost are damages incurred, in the form of stolen assets, when these costly security measures fail to successfully defend against attacks by malicious actors. According to a study conducted by The Atlantic Council, in collaboration with Zurich Insurance Group and the Pardee Center for International Futures, this spending is expected to rise rapidly and reach a staggering cost of $160 Trillion in the year 2030 (*17*). Such an immense cost should be accompanied by an equally rising benefit/advantage provided by the associated cybersecurity measures. The aforementioned study also concluded, however, that no study has successfully measured the costs of cybersecurity problems compared to the economic benefits of being [digitally] connected. Given the severity of this expected rise in cost of security, a diminishing benefit to cost ratio is a reasonable possibility. Bitcoin offers a likely solution to this crippling increase in cost of security.

Bitcoins security is a result of the myriad of nodes around the world mining it. This mining requires electricity which incurs an associated cost. According to Digiconomist, a web application designed to estimate the amount of power and related cost with worldwide Bitcoin mining, the cost as of May, 2018 is roughly $3.3 billion (*18*). Concurrently, the annualized global revenue in the form of Bitcoin rewards accompanying this mining generates approximately $7.6 billion. This complementary reward is an ephemeral phenomenon, however, and will eventually disappear and be replaced by transaction fees as opposed to the minting of new coins once all the possible Bitcoin have been mined as a result of the circulating supply reaching the maximum of 21 million coins. As a result, in the future there will be a net cost for securing the Bitcoin blockchain. Bitcoin mining power usage and cost is expected to plateau and at most rise linearly unlike the exponential rise in cost of traditional cybersecurity measures. Unlike the cost to benefit ration of traditional cybersecurity, analyzing this same ratio associated with securing the wealth proportional Bitcoins market capitalization is a straightforward calculation. The result of this calculation currently holds little meaning due to the volatility of Bitcoin. Assuming these trends of traditional security costs and future Bitcoin mining costs are accurate, it can be concluded that Bitcoins model for securing wealth and assets is more cost-effective than traditional methods.

Future efforts in this space should focus on the development of valuation methods that are more generalizable.

6 The Future Role of Modeling and Simulation

In order to understand what factors may play a role in the creation of a successful PB, one could begin to apply agent-based modeling approaches derived from complexity science. By modeling a blockchain deployed in a decentralized context, perhaps incorporating the competitive advantages natural resources may confer, we can begin to draw conclusions about potential human-participation patterns that may emerge from a given design point. There are many potential dimensions of design worthy of exploration in this space. Individual and group incentives, consensus algorithms and validation mechanisms, social network effects, and multi-PB environments are all potential targets of future modeling efforts.

7 Conclusion

The emergent properties of public blockchain technology are not well understood, and they do not lend themselves to easy quantification. By viewing public blockchain technology as a complex adaptive system, we can begin to have the discussions required to engineer and describe the value provided by this class of systems in a truly holistic manner. Developing complex adaptive systems-based models of public blockchain designs that take into account the incentives of the humans to participate, especially given the vast array of existing alternatives available, may become necessary in order for public blockchain engineering efforts to achieve their goals.

Acknowledgments

The authors wish to specifically acknowledge Charles J. McCarthy for his support of this effort. The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions or viewpoints expressed by the author.

References

1. M. D. Norman, M. T. Koehler, and R. Pitsko, "Applied complexity science: Enabling emergence through heuristics and simulations," *Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach*, pp. 201–226, 2018.

2. M. D. Norman, "Complex systems engineering in a federal IT environment: Lessons learned from traditional enterprise-scale system design and change," in *2015 9th Annual IEEE International Systems Conference (SysCon)*, pp. 33–36, IEEE, 2015.

3. G. H. Lewes, *Problems of Life and Mind*. Houghton, Osgood, 1879, 1873.

4. Rapid7, "Common types of cybersecurity attacks: A look inside the attacker's toolkit," 2016.

5. Investopedia, "51% attack," 2018.

6. E. Foundation, "Ethash consensus algorithm: An example of the necessary, standardized set of rules," 2017.

7. Y. Bar-Yam, *Dynamics of complex systems*, vol. 213. Addison-Wesley Reading, MA, 1997.

8. Investopedia, "What is the difference between economic value and market value?," 2018.
9. F. S. Jeff Kauflin, "The fast-growing job with a huge skills gap: Cyber security," 2017.
10. D. Bonderud, "Demand for cybersecurity talent soars, study finds," 2018.
11. M. Iansiti and K. R. Lakhani, "The truth about blockchain," 2017.
12. Investopedia, "Liquidity," 2018.
13. A. de Tocqueville, *L'Ancien Rgime et la Rvolution*, vol. 3. 1850.
14. S. Raval, "Decentralized applications," 2018.
15. M. Orcutt, "The tricky art (and emerging science) of valuing crypto-assets," 2018.
16. Statista, "Spending on cybersecurity in the united states from 2010 to 2018 (in billion u.s. dollars)," 2018.
17. Z. A. I. Company, "Atlantic council / zurich insurance report finds the global benefits of cyber connectivity expected to outweigh costs by $160 trillion through 2030," 2015.
18. Digiconomist, "Bitcoin energy consumption index," 2018.

Social Valuation Dynamics for the Blockchain and Cryptocurrency Era

Lynette Shaw,[1*]

[1]Center for the Study of Complex Systems, University of Michigan

*To whom correspondence should be addressed; E-mail: shawla@umich.edu

In the wake of the Great Recession, a monetary object born of a novel combination of cryptography and anti-centralization politics was introduced to the world: Bitcoin. In less than a decade since, the collective valuation of cryptocurrencies has risen into the billions (USD) and many hundreds of millions (USD) more has been invested into the development of blockchain technology applications. Accompanying this rise have been many questions concerning what this new technology constitutes a case of and how it has come to have value in the absence of state or commodity backing. This paper overviews prior research (*1*) that considers how we might shed light on these questions by applying computational social science methods and agent-based modeling to better understand these dynamics of collective definition and social valuation that have driven this rise of cryptocurrency and blockchain technology.

Introduction

The ascent of cryptocurrencies and blockchain technology brought to the forefront a classic economic and social problem from where does value originate? Unlike monies backed by precious metals, Bitcoin and the majority of alt-coins which have followed since its inception are not backed by commodities that are considered to have intrinsic worth (contra (*2*)). In purposeful contrast to traditional fiat monies, the majority of new currencies and tokens also lack any involvement by states or formal governing bodies (contra (*3*)). Even cast in technological rather than monetary terms, blockchain participation still often faces a fundamental bootstrapping issue wherein the value of joining the network depends fundamentally on the number of others who have already joined it (i.e. network effects). While it might be obvious why individuals would choose to join a mature network of this sort, it is not immediately evident why individuals would make the choice to opt into an early network in sufficient number to get it past critical adoption thresholds required for its success. In spite of their inability to be readily explained via such established models, however, in the less than decade since their inception, the total market capitalization of all cryptocurrencies stands in the hundreds of billions (USD).

This foundational problem of how, even in the absence of innate individual utilities or centralized institutions, many individuals can come to collectively regard an object as having worth is a classically sociological topic and speaks directly to a central object of study within the field, that of the social construction of value (*4, 5*). A simple explanation of social construction is offered via the so-called Thomas Theorem (*6*) which states, If men define situations as real, they are real in their consequences. Applications of social construction models are often most strongly associated with the study of race (*7*) and gender (*8, 9*), but there are numerous examples of excellent classical and contemporary research in economic sociology and organizational studies that have successfully mobilized social constructionist perspectives toward the end of better understanding a variety of economic phenomena including money (*10, 11*), market valuations of stocks and companies (*12*), and pricing environmental damage (*13*). As the dawning of the cryptocurrency and blockchain era brings the fundamentally social processes underlying valuation back out into focus, the necessity of applying, extending, and formalizing the insights of such constructionist traditions becomes increasingly apparent.

This paper overviews prior research (*1*) undertaken in this regard in the hopes of inspiring further work on how the social construction of value and its associated dynamics apply to cryptocurrency and blockchain technology. The first part offers a brief review of empirical work which applies models of the interplay between categorization and valuation developed in the context of organizational studies (e.g. (*12, 14, 15*) to unpack the processes of collective definition which drove Bitcoins initial trajectory of adoption and valuation and ultimately led to the emergence of the more general blockchain technology market category. The second part then overviews work on the development of a formalized theory of socially constructed value that is amenable to systematic elaboration via computational (i.e. agent-based) modeling and can account for a range of phenomena relevant to cryptocurrency and blockchain valuation dynamics including the emergence of stable value conventions, the influence of initial conditions and early confident actors on social valuation, and the ability of transient bubbles or panics to produce a time-varying ratcheting effect in an objects conventional valuation.

Categorization and the Evolution from Radical Money to Blockchain Technology

A foundational understanding established in the past two decades of work within organizational studies is that valuation depends fundamentally upon categorization (*12, 14–17*). As social groups define which category an object should be classified as belonging to, they also establish the theories of value (*18*) that will be applied in assessments of its economic worth. Within the highly decentralized context of cryptocurrencys history of development and adoption, there have been significant questions surrounding how it should be classified and its value understood. Is it money? A payment system or commodity? The future of digital commerce, a speculative investment, a criminal currency, or a scam? A disruptive new technology that will revolutionize finance and business? A challenge to central governments and existing financial institutions or the pet project of a group of political fringe actors?

A main contention of previous research on this topic (*1*) is that at different times and to different constituencies, cryptocurrency has managed to fall into all of these categories. In the context of established markets, this sort of multivalent identity (*19*) might be expected to lead to a so-called illegitimacy discount (*12*) wherein cryptocurrencys failure to be cleanly defined as belonging to one category or another would lead to confusion in how to assess its economic worth and thus by default, lead to lowered market valuations of it. As more recent categorization research has established, however, in the context of nascent technologies and emerging markets, this level of ambiguity in identity can be a benefit (*14, 20, 21*). As argued in this prior work (*1*), this proved to be the case for cryptocurrency. More specifically, this research demonstrates how this multivocal (*22*) quality of cryptocurrency directly enabled its trajectory of adoption and development from its inception through 2015 by enabling different groups to find different ways of interpreting its potential uses and thus construct different reasons to adopt it and opt into contributing to its development. Beginning from the ideological motivations that spurred Bitcoins initial creation as a politically subversive alternative money, through the community ethos that drove its early development as a open source project, to its translation into a medium of exchange for illicit online activities, and further into recent approaches to cryptocurrency as alternatively a digital asset, a payment system, or a new disruptive technology for finance and business applications, the story of cryptocurrencys rise is one that has been marked by an expanding constituency of audiences who have found their own reasons for buying in.

The diversity of the ecosystem that has emerged around cryptocurrency and blockchain technology might be seen as a direct consequence of the highly decentralized context which facilitated this involvement of so many different types of groups. In a fulfillment of the aspirations of many decentralization advocates, the involvement of everyone from political radicals to major players in the existing finance system in the exploration (*23*) of what cryptocurrency might be defined as acts as a testament to the democratizing potentials of decentralization. As also shown in this prior work, however, mounting pressures to narrow down cryptocurrencys identity (i.e. move from an exploration of what it might be into an exploitation (*23*) of just a few facets of it) has resulted in increasingly significant tensions within this ecosystem. Using a combination of venture capital funding data, Google search trends, and automated content analysis of over 7,500 news articles on Bitcoin and blockchain related topics from 2011 2016, this analysis demonstrated how the influence of later arriving but well-resourced audiences dramatically shifted the collective definition of cryptocurrency away from conceiving of it as a basis of alternative payment systems and toward an understanding of it more in terms of a blockchain technology which might be applied toward the end of increasing the efficiency of existing business and financial applications (see Figure 1 and Table 1 for an example of these analyses).

By examining the processes of collective definition and categorical contention responsible for the evolution of cryptocurrency from a radical political money into a new technology for establishment finance and banking sectors, the ultimate aim of this work has been to invite deeper consideration of the fundamentally social dynamics which drive valuation in emergent arenas and to direct attention to the specific manner in which they manifest in highly decentralized contexts.

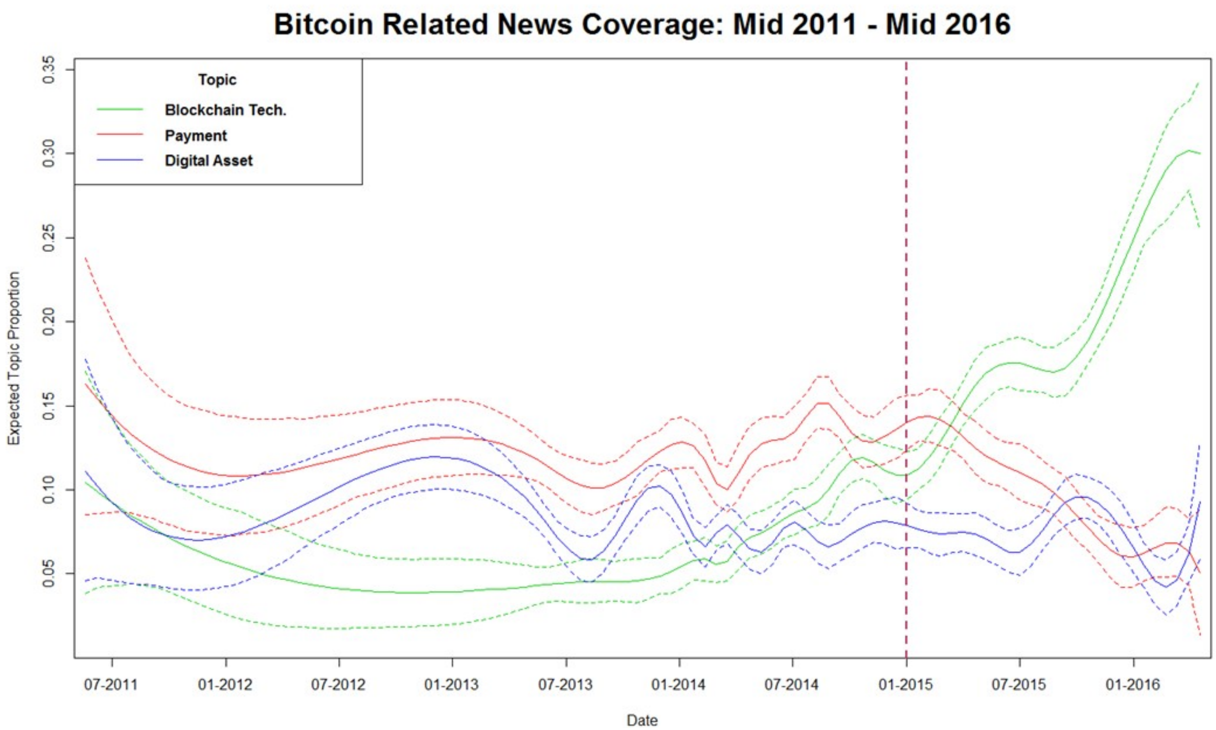


Figure 1: Estimated topic proportions for selected topics, mid-2011 through early 2016, for entire corpus (7,903 documents, 6 sources). Vertical dashed line indicates the beginning of the quarter (Q1 2015) which saw major shifts of venture capital funding into business and financial applications.

Topic Number	Topic Label	Top 15 Words (by FREX score)
10	*Blockchain Technology*	blockchain, technolog, project, ledger, rippl, distribut, ethereum, smart, contract, applic, innov, platform, infrastructur, decentr, build
7	*Payments*	card, merchant, coinbas, payment, mobil, accept, fee, appl, app, paypal, option, credit, pay, bitpay, processor
3	*Digital Asset*	price, china, trade, chines, volum, trader, percent, investor, invest, stock, index, market, liquid, volatil, winklevoss

Table 1: *Blockchain Technology*, *Payments*, and *Digital Asset* topics with top 15 words

A Computational Model for the Social Construction of Value

In addition to technological disruptions, the rise of cryptocurrency has also brought with it any number of conceptual ones as well. Of these, one of the most important has been to standard models of how money acquires and holds value. Absent a backing such as gold or the state, the question of how digital currencies have attained significant real-world valuation in the over nine years since its inception remains a puzzle to many standard economic models. Many have argued that a chief reason for this inability of prevailing economic models to satisfyingly explain valuation situations like the one surrounding cryptocurrency and blockchain adoption stems from the assumptions standard economic models make on the independence of value from social processes (*11, 24–26*). These critiques assert that if one begins from a standpoint of treating value as existing inherently in objects, then it is very difficult to explain how socially based objects like cryptocurrency or money become something out of nothing in terms of their collective valuation. In contrast to these traditional economic conceptions of value is the large body of work in sociology that has delved directly into the influence of social construction processes on valuation (*4, 5, 10, 27*). Though varying in their substantive applications and theoretical emphasis, these sociological models share a foundational focus on how the interdependency of individuals valuation processes makes it possible for value to arise from social interactions alone.

Prior theoretical work (*1*) uses cryptocurrency as a motivating case to translate these traditionally qualitative models of socially constructed value into a formalizable, computational model that allows for a more systematic and rigorous exploration of how something is able to arise out of seeming nothing vis--vis social interaction. It proceeds on the basis of identifying a significant point of commonality between the realist (*5*) approach to value favored in economic models that asserts the existence of intrinsic value and constructionist (*5*) or conventionalist (*10*) approaches that dominate in more sociological treatments which assert that the value of an object is ultimately just the value that a social group collectively assigns to it. As divergent as these two perspectives are, this prior work contends that underlying both is a conception of individual valuation as being essentially a process of individual learning under uncertainty. From this basis, the model then contends that the main difference between these two perspectives reduces to the different levels of dependency individuals have on social vs. non-social feedbacks in the learning/valuation process.

Given this foundation, this theorization makes a conceptual bridge between valuation and formal models of parameter estimation that rely on updating an existing model of reality with information from observed data (i.e. learning). The key link between the conceptual and computational model is to liken an individuals attempt to arrive at the correct estimation of an objects value to an agent attempting to use the data it observes to estimate an unknown parameter, θ, underlying a Bernoulli data generation process via a recursive Bayesian updating (Eq. 1):

$$p(\theta|y_t) = \frac{\Gamma(a+b+n)}{\Gamma(a+y_t)\Gamma(b+n-y_t)}\theta^{a+y_t-1}(1-\theta)^{b+n-y_t-1} \tag{1}$$

Where a and b are the shape parameters of the agents current prior at time t, n is the number of agents that were sampled and y_t is the number of the agents sampled that turn who were observed in the up state. To capture the state of initial uncertainty, all learning agents also begin the learning-valuation process with flat, weak priors (i.e.$a_0 = b_0 = 1$) unless otherwise specified.

Using this agent as its basis, this work analyzes a series of computational models that investigate how collective processes of parameter estimation develop under different information feedback scenarios. Of primary interest are pure non-social (i.e. realist) scenarios where agents use only information from a special set of fixed, non-social agents whose behavior is determined by some true θ_{fixed} that they all share, pure social (i.e. constructionist) scenarios wherein no θ_{fixed} exists and learning agents observe only other learning agents whose behavior is based on their current estimates for θ, and mixed scenarios in which agents observe a combination of social and non-social feedbacks. Using these results as a foundational baseline, this prior work then develops a series of substantively interesting variants to isolate the effects of initial conditions, the presence of initially confident actors, and the effect of exogenously induced bubbles and panics on emergent collective valuations. The results yielded by this modeling approach offers a rich set of insights into the emergent dynamics of social valuation. Space limitations prohibit a full exploration, but some of the results most relevant to cryptocurrency and blockchain buy-in are worth noting.

The most foundational finding is that even in the absence of any initially true θ_{fixed} pure social systems readily converge upon stable, self-reinforcing conventional values for θ and do so on a timescale comparable to purely non-social systems (see Fig. 2). In addition to demonstrating that purely social systems do converge, this simulation approach also facilitates an exploration on what conventional values individual systems ultimately converged upon (see Fig. 3).

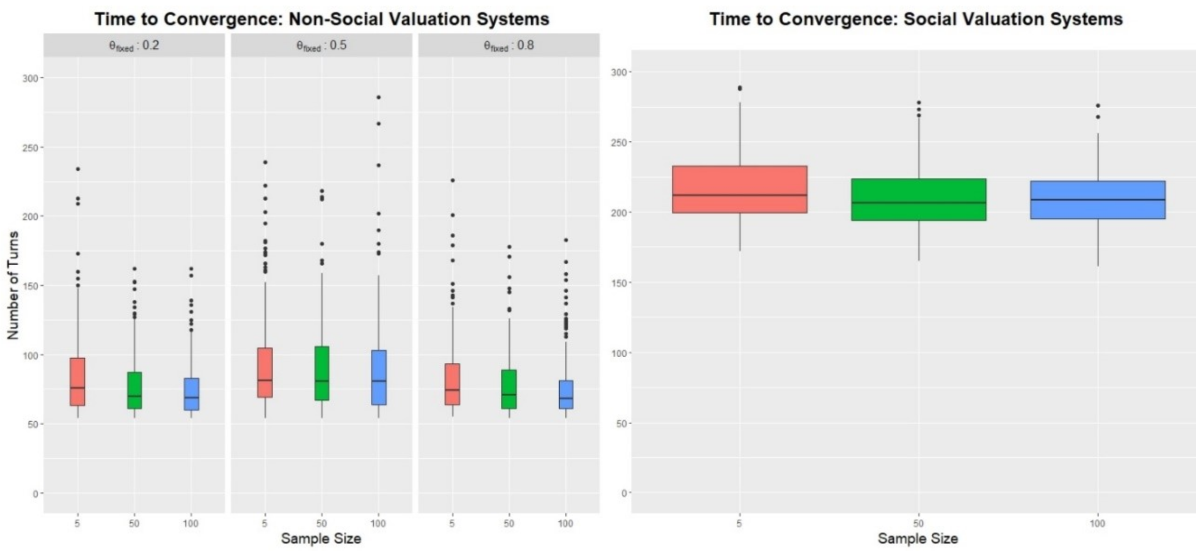


Figure 2: Distributions of time to convergence for series of social valuation simulation runs (200 runs per scenario).

The primary takeaway for the cryptocurrency and blockchain context of this demonstration of how something can readily arise out of nothing via the process of valuation itself is an affirmation that even in the absence of a state or commodity backing, self-reinforcing conventional valuation of objects should be treated as being fundamentally stable in nature. However, the variability of where individual runs of pure social systems ultimately settled in their valuation convention also points toward how deeply contingent and path dependent such social valuation dynamics are in their evolution. What this variability fundamentally indicates is how small accidents or alterations in individual behavior early on can have long term consequences for the eventual state.

One exploration of exactly this issue is available through a substantive variation in the baseline model which considers what happens when systems are exogenously forced in the first turn of interaction to look as if many agents do or not value the object under consideration (as indicated through the proportion of agents initialized in the up state see Fig. 4 for results). To note, in these systems, all agents began in the same conditions of initial uncertainty as before with the only outside intervention being the outward state they were initialized in. Nonetheless, as demonstrated in Fig. 4, this single turn of initial interaction is sufficient to strongly drive where the ultimate value convention of the system settles. Another exploration considers the effect of an individual or small group of individuals who begin with a strong, though technically incorrect idea, that the object under consideration is highly valued by the social group (i.e. some subset of agents begin with a strong prior of θ= .99 - see Fig. 5 for results). In both

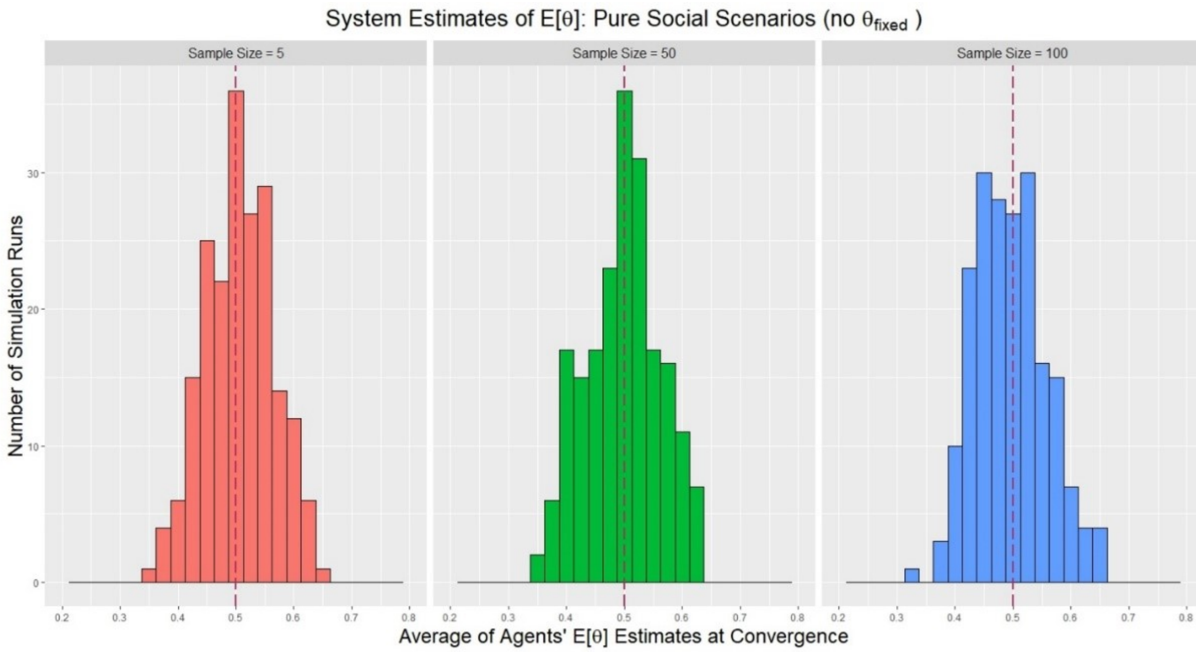


Figure 3: Distribution of what pure social learning systems converged upon as their conventional value of θ

scenarios, the effects of initial conditions or confident, even if wrong, actors at the outset of the social valuation processes is clear. When it comes to the case of social valuation as it pertains to blockchain contexts, these results indicate ultimately that early appearances even if not truly indicative of the systems real state of valuation initially can have a profound impact on where the real conventional value of on object ultimately end up.

A final substantive scenario which has a particularly direct bearing on the valuation of cryptocurrencies concerns the interplay of conventional valuation and the sorts of animal spirits (*28*) and irrational emotional contagions that are known to produce both panics and bubbles in markets. As the set of foundational results attest to, socially constructed value may involve high levels of initial contingency in its emergence but this does not in anyway entail that it is innately unstable. Consequently, criticisms that the valuation of Bitcoin or other cryptocurrencies or digital tokens are nothing more than a long-term bubble that is equivalent to the purported tulip mania of 17th century Holland are unfounded. Nevertheless, it is hard to deny that the historical development of cryptocurrency has been marked by periods of both outsized collective effervescence and fear which have had profound impacts on their market valuation.

The long term impacts of such periods on conventional valuation can actually be captured through a substantive variation in the model that incorporates brief periods in which some subpopulation of agents temporarily behave as if their current, learning-based expected value of θ is inflated by some factor (i.e. 25% of the systems population of agents change to the up state with a probability of $E[\theta] + \alpha$ for 50 turns of interaction see Fig. 6 for average trajectory runs).

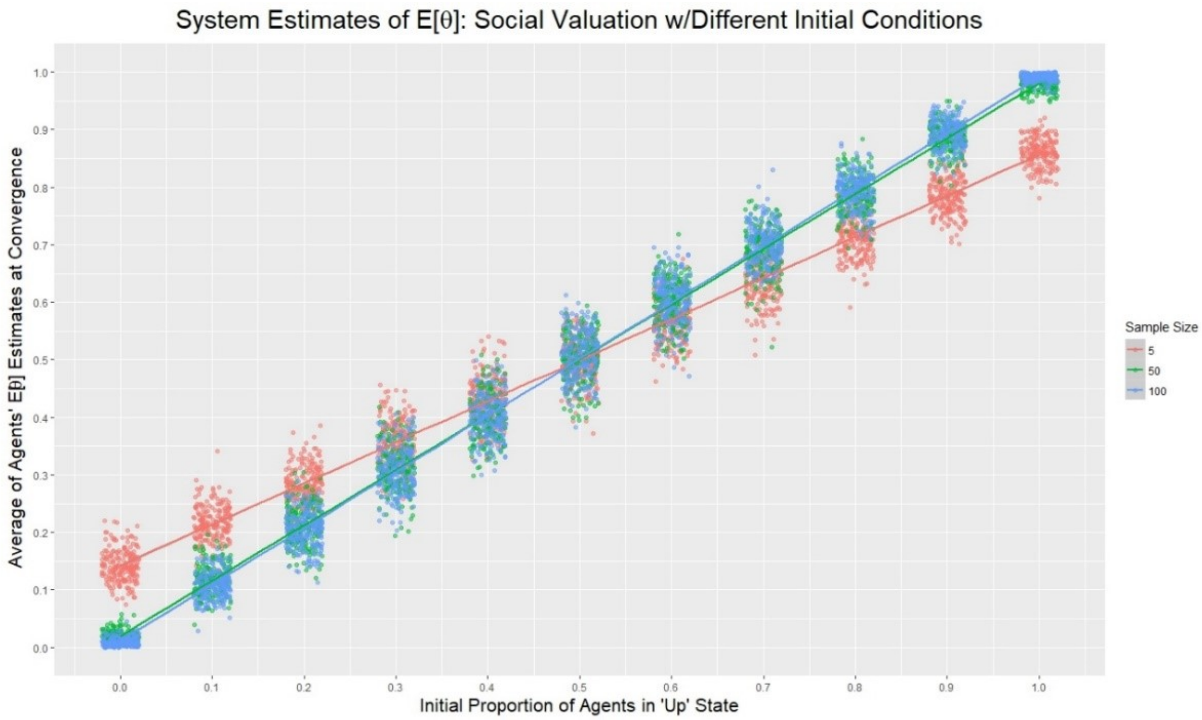


Figure 4: Distribution of what pure social learning systems converged upon as their conventional value of θ versus what proportion of agents were initialized in an up state

These theoretical results demonstrate how in the case of pure social valuation, volatility in terms of bubbles and crashes in value are still produced in the presence of emotional contagions that lead to temporary, mass irrationalities in collective valuation. Significantly, however, in cases of pure social valuation, something that might be called the fossilized remnants of earlier animal spirits can continue to persist via a ratcheting effect they produce in where the long-term conventional value of the system lands, as seen in Fig. 6 where the stabilized value crashes post-bubble, but at a level higher than the previously established convention. Furthermore, the influence of this ratcheting effect can be seen to diminish over time, at least in the closed systems considered here, with later bubbles have less of an impact than earlier ones. The degree to which this set of findings resemble the qualitative facts of Bitcoins and other cryptocurrencies market value through time offers encouragement for the appropriateness of applying a social constructionist model to them.

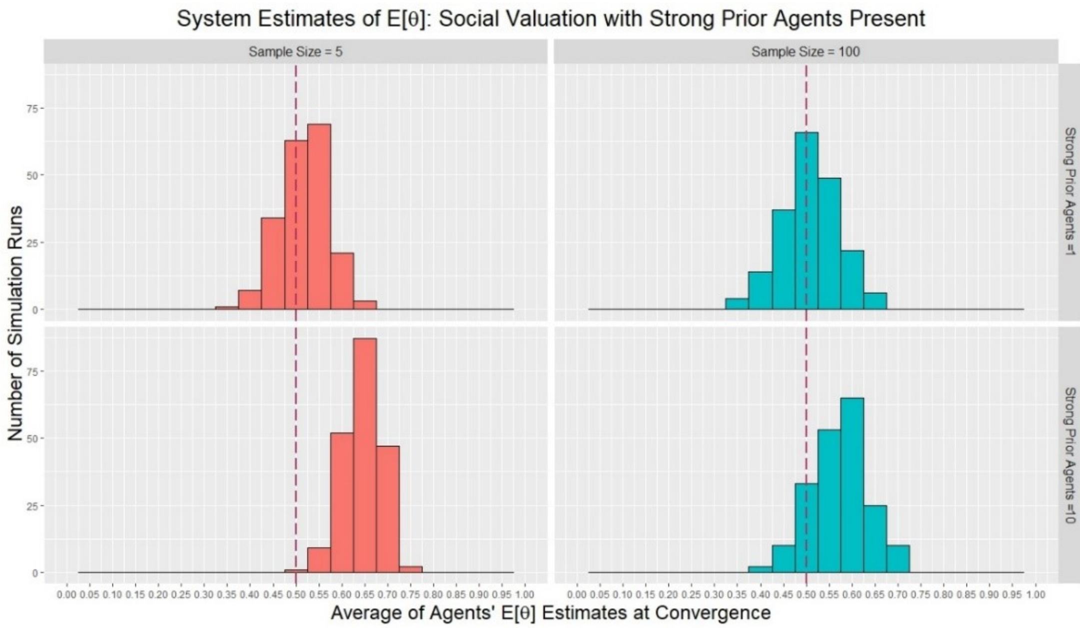


Figure 5: Distribution of what pure social learning systems converged upon as their conventional value of θ in the presence of either 1 or 10 agents who began with strong priors

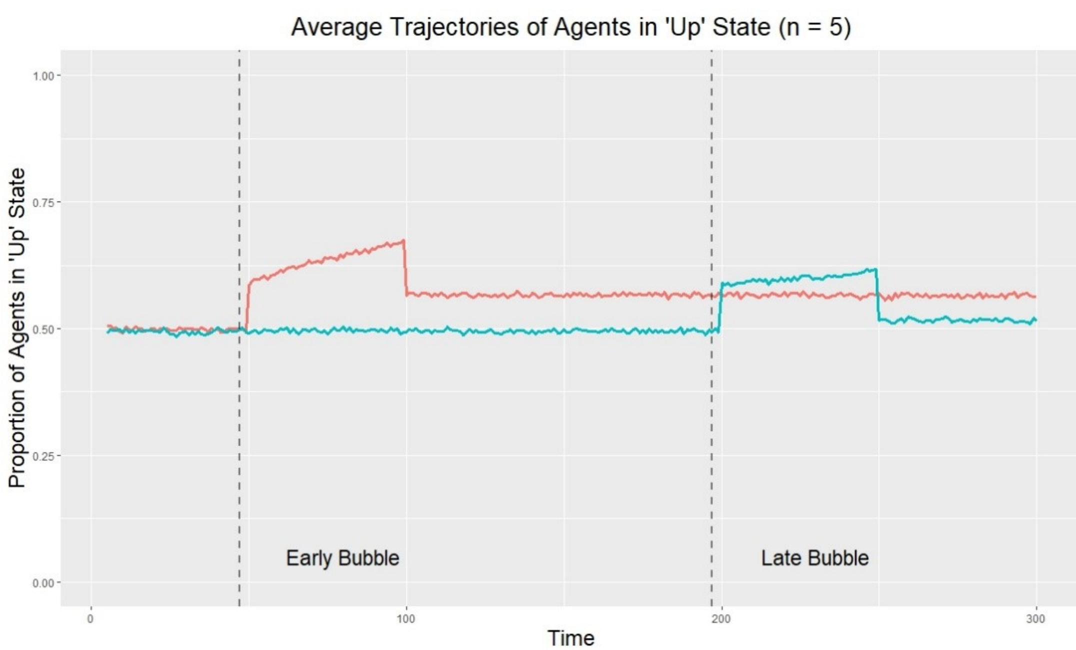


Figure 6: Average trajectory of 200 runs under each of bubble scenario

Conclusion

The rise of cryptocurrency and blockchain technology has brought with it a host of implications, the majority of which we likely have yet to understand. Among these are not only practical implications for the electronic exchange of value and the development of a world of decentralized applications, but also conceptual implications for how we understand the nature of money and value itself. Oftentimes, these subjects are so deeply taken-for-granted in daily life, that we have little cause to even consciously think about them. With the blockchain era now upon us, however, we have little choice but to revisit our existing assumptions and established models and interrogate how they hold up to this radically new context. The foregoing has offered a very brief overview of existing work that demonstrates the need to engage with the fundamentally social the fact that this exploration (*23*) of what cryptocurrency might be defined as and used for dynamics driving economic valuation in the cryptocurrency and blockchain technology contexts with the hope that these analyses will encourage others to pursue further research at this rich, and highly generative intersection.

References

1. L. Shaw, The Meanings of New Money: Social Constructions of Value in the Rise of Digital Currencies, Doctoral dissertation, University of Washington (2016).

2. K. Menger, *The Economic Journal* **2**, 239 (1892).

3. G. F. Knapp, *The State Theory of Money* (Augustus M. Kelley, New York, 1924).

4. M. Lamont, *Annual Review of Sociology* **38**, 201 (2012).

5. E. W. Zuckerman, *Annual Review of Sociology* **38**, 223 (2012).

6. W. I. Thomas, D. S. Thomas, *The Child in America: Behavior Problems and Programs* (Knopf, New York, 1928).

7. I. F. Haney-Lopez, *Harvard Civil Rights-Civil Liberties Law Review* **29**, 1 (1994).

8. J. Butler, *Gender Trouble: Feminism and the Subversion of Identity* (Routledge, New York, 1990).

9. C. West, D. Zimmerman, *Gender and Society* **1**, 125 (1987).

10. A. Orléan, *The Empire of Value: a New Foundation for Economics* (MIT Press, Cambridge Mass., 2014).

11. V. Zelizer, *The Social Meanings of Money* (Basic Books, New York, 1994).

12. E. W. Zuckerman, *American Journal of Sociology* **104**, 1398 (1999).

13. M. Fourcade, *American Journal of Sociology* **116**, 1721 (2011).

14. R. Durand, M. Khaire, *Journal of Management* **43**, 87 (2017).

15. M. Hannan, L. Pólos, G. Carroll, *Logics of Organization Theory: Audiences, Codes, and Ecologies* (Princeton University Press, Princeton, 2012).

16. M. T. Hannan, *Annual Review of Sociology* **36**, 159 (2010).

17. E. W. Zuckerman, T.-Y. Kim, K. Ukanwa, J. von Rittmann, *American Journal of Sociology* **108**, 1018 (2003).

18. E. W. Zuckerman, *From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads*, R. Durand, N. Granqvist, A. Tyllström, eds. (Emerald Publishing Limited, 2017), pp. 31 – 68.

19. G. Hsu, M. T. Hannan, Ö. Koçak, *American Sociological Review* **74**, 150 (2009).

20. N. Granqvist, S. Grodal, J. Woolley, *Organization Science* **24**, 395 (2013).

21. E. G. Pontikes, *Administrative Science Quarterly* **57**, 81 (2012).

22. J. F. Padgett, C. K. Ansell **98**, 1259 (1993).

23. J. G. March, *Organization Science* **2**, 71 (1991).

24. G. Ingham, *The Nature of Money* (Polity Press, Cambridge, 2004).

25. A. Orléan, *Journal of Economic Behavior and Organization* **28**, 257 (1995).

26. G. Simmel, *The Philosophy of Money* (Routledge and Kegan Paul, Ltd., Boston, 1978).

27. L. Boltanski, L. Thévenot, *On Justification: Economies of Worth* (Princeton University Press, Princeton, 2006).

28. J. M. Keynes, *The general theory of employment, interest and money* (Harcourt, New York, 1936).

Asymmetric Trust and Causal Reasoning in Blockchain-based AIs

Percy Venegas[1]

[1] Economy Monitor

Abstract

We use genetic programming evolved networks, vector fields, and signal processing to study time varying-exposures where trust is implied (e.g. a conversion event from attention flow to financial commitment). The datasets are behavioral finance time series (from on-chain data, such as fees, and off-chain data, such as clickstreams), which we use to elaborate on various complexity metrics of causality, through the creation parametric network graphs. We discuss the related methods and applications and conclude with the notion of social memory irreversibility and value by memory as useful constructs that take advantage of the natural fact of the existence of trust asymmetries, that can be operationalized by embedded AIs that use distributed ledgers both as the substrate of their intelligence and as social computers. By being context-aware, those intelligent agents are able to intervene in problematic stressors and contribute to minimizing network fragility.

Keywords: systemic risk, behavioral finance, economic complexity, evolutionary computation, computational trust, the blockchain, cryptocurrencies, market microstructure, reality mining.
JEL Classification: G02, F63, B17, C53, C58

There is none *deceived* but he that *trusts —Benjamin Franklin*

Introduction

Given the current state of knowledge, it is relatively easy to have artificial intelligent agents to find patterns and to formulate predictions following some objective criteria. But to get become useful in comparison with human intelligence, it is crucial that those agents are able to ask: Why? Posing the question is an exercise on causal reasoning, a realization of the awareness of cause and effect. It is also a matter beyond logic formalisms— the interface matters. That is because once the intelligent agent is able to ask why?, it will be also in a position to ask counterfactual questions, such as how an intervention may change the output, or even the causal relationship itself.[1] This is particularly important in the case of blockchain-based AIs because directionality matters (causation works in one way) and because the substrate of the AI is a fully or partially immutable ledger.

One may attempt to reduce the proposed solution to the problem to the application of well-known methods, such as Bayesian networks; since blockchain systems are deterministic and it is desirable that the constructs operating on top of them (e.g. smart contracts) behave with some degree of predictability, establishing an appropriate intuition of the priors by access to on-chain data and some sort of data integrity-verified artifacts (e.g oracles) may at least partially achieve this purpose. However, such a simplistic approach would deprive the AI of context: the world is not presented to us as a data feed, but rather as a dynamical experience in which the embeddedness into a social context[2] dictates the response, especially in the forming stages of intelligence. Particularly, the repeated use of metaphors such "circles of trust" in industry and personal relations hints at the tendency that humans have to at least implicitly compute similarity metrics (to define the boundaries of that circle or space) and to elaborate mechanisms to detect violations to certain social laws and descriptive models of economic behavior.[3] In a way, to engage in the world we require that other agents are "trust verified".

Trust is fundamental to the human experience, yet it is little understood. But AI, web analytics, and blockchain technology have come to change that. For the first time in history, we have real-time data to map how attention flows, and understand how people actually assess risk and commit resources. And with AI we can augment our own intellect,[4] to understand complex socio-technical systems. In some cases, this is a full departure from the prevailing economic theories, that were developed using experiments conducted on small groups, incomplete and delayed macroeconomic data, or theoretical models which are completely dissociated from reality. There is no reason why economics and the social sciences have to be called "soft" anymore; there is no such thing as hard and soft sciences, a scientist should always operate in the realm of facts and quantitative evidence, otherwise, he is only a commentator.

Computationally, biologically, and socially, humans "need" to trust.[5] And even when dealing with "trust-less" systems such as distributed ledger technology, everyone (including AIs) need to trust "on the design". This work is concerned with the role of trust asymmetry[6] in the causal reasoning process of blockchain-based AIs. We will approach it from the perspective of the machine, of the method and apparatus needed for the intelligent agent to make consequential decisions in the world: is there an asymmetry, then why? And, where and when trust asymmetry breaks?

Related Work

Integrating social information with traditional network layers

A blockchain-based AI becomes aware of the environment via off-chain data. To conceptualize this using the OSI model as a reference, we use a cross-layer stack (layer 7 and layer 6). In practice, the financial activity logged in the blockchain is the expression (e.g. a conversion event) of the consumption, flow of attention, and commitment of resources in adjacent networks such as the web and mesh IoT. For instance, when a web browser creates a request (e.g. GET / HTTP) a Java-based application could log the hits, detect the device type (mobile or desktop), and other features included in the user-agent header. Also by looking down in the stack to the routing level, ISP data is used to obtain geographical origin, redirect path and destination. This sort of "alternative data" becomes especially useful when studying permissioned and semi-centralized networks, since not all data is publicly available and suitable proxies are required.

Ripple

Formally, Ripple is not a blockchain, but a common ledger based on a proprietary technology to cater to the privacy needs of the banking industry —therefore, some transactional and network activity details are unavailable to the public. We tracked daily usage for the 100 most popular services among prospect Ripple users over the course of 18 months (548 days in total). We use daily prices as target variable since a general audience-prospect user will be inclined to look at the daily prices, while professional investors usually focus on daily returns. The services included those directly related to cross-border payments operations (e.g. gateways) and other peripheral to the economy, including price information services, wallets, and the like. We investigate the long-term market structure, specifically the demand signals from that segment of newcomers. We started with one hundred services, and after many rounds of elimination making different formulas compete with each other for accuracy (using symbolic regression), the simplest and most meaningful relationship expresses price as a function of two constant values and the demand for the services of a particular exchange. In other words, the simpler predictive model to provide any insight traces back the rise of Ripple among this segment to the popularity of one of the prominent exchanges (of the centralized type) that listed the coin. One such predictive model can be expressed as $Price = a * exchange_1 - b$

A more complex model has the form $Price = a + b * sma(wallet\_users, 21) * sma(wallet\_users, 37) + c * wallet\_public^2 - d * exchange - e * gateway * sma(wallet\_users, 21) * sma(wallet\_users, 37)$

According to this, from May 2017 to December the use of a particular *wallet* created support levels of 21 and 37 days (using simple moving averages) and it had the bigger impact of all variables discovered (increasing the use of this wallet has a positive impact on price 100% of the times). This means that the usage of this service serves as a "canary in the mine", i.e a prolonged decrease in usage (being all conditions equal, such as not having a similar alternative replacing the use of the wallet) will indicate weakening demand fundamentals.

The first negative term is a *gateway*, which in Ripple's architecture means "businesses that provide a way for money and other forms of value to move in and out of the XRP Ledger". This particular issuing gateway supports both Yuan/XRP and Yuan/Stellar pairs and provides services mainly in China, and to a lesser extent in Japan and the US. This same gateway's popularity increased following the crackdown on bitcoin exchanges in China.

The second negative term is a Chinese *exchange*, of the centralized type. That is, as demand for the exchange of CNY/digital asset increases, this may be exercising some downward pressure on price. The negative effect is slightly larger when people use the centralized exchange, rather than a Ripple gateway. This may suggest that some operatives turned to Ripple as a haven when the Bitcoin exchanges where hit, although prices are still susceptible to movements of XRP assets in and out of the economy via a gateway. But strong demand from the middle market supersedes the fears of those uncomfortable with all-time-highs, and this is why the usage of the Ripple wallet exploded on Dec 13–14th, in tandem with the spike in price volatility of XRP —alongside with attractive dynamics of the BTC/XRP pair. It is also important to note what the AI does not see: the lack of any oscillatory term in the formula $(\sin, \cos)$ hints at the lack of strong regularities (periodicities) during the eight months of the analysis. The other valuable observation here is that in the absence of access to many of the Ripple ledger statistics that will normally allow identifying large holders, the usage of the wallet allows to single out the mid-market as a force driving the market (and this wallet service is predominantly accessed from the US).

Time series analysis and prediction

The application of genetic programming to the study of behavior and causation in cryptocurrency markets is not only an analytic artifact, but it is fundamentally aligned to the nature of the problem.

Taleb and Douady[7] explain that the natural selection of an evolutionary process is particularly antifragile since a more volatile environment increases the survival rate of robust species and eliminates those whose superiority over other species is highly dependent on environmental parameters. In the context of cryptocurrencies, we could use temporal correlations in blockchain traffic to gauge the response of a given object (e.g. fees) to the volatility of an external stressor that affects it, but another approach is to simply study the response of the market (as measured by a common risk metric, such as volatility) to the actual market behavior (with the consumption of services and information measured by HTTPs requests, including endogenous variables such as the activity at the customer service channels of the wallet and the exchange itself, mining pools, mining profitability feeds, and so on; in this toy example we use just a subset of those variables). The driving variables are modeled using symbolic model ensembles, as in Figure 1, which is based in daily time series for the period of November 1st, 2016 to May 9th, 2018.

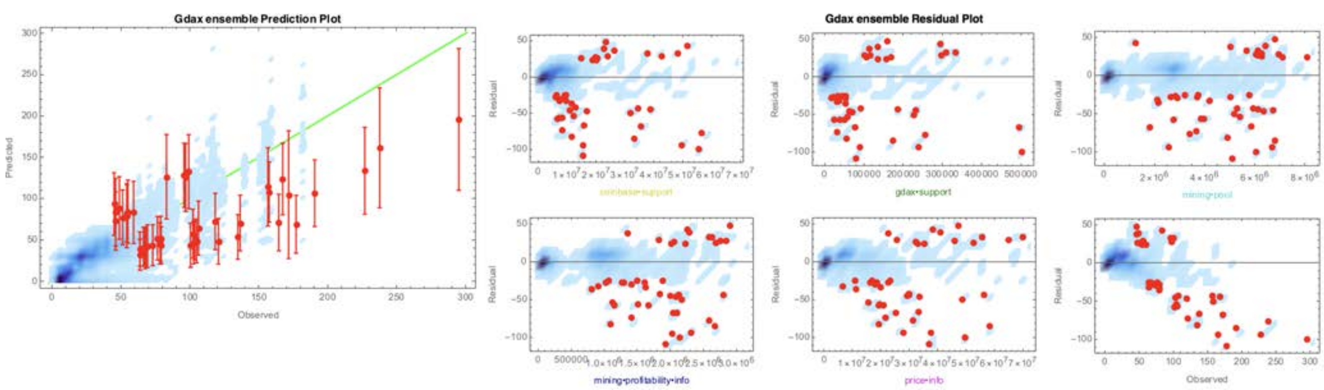


Figure 1: Volatility of BTC price in cryptocurrency exchange and environmental (demand) signals

The nonlinearity in a source of stress is necessarily associated with fragility. This is perhaps why low-quality coins fail —marketing activity is an environmental variable only, while actual installed capacity and operational infrastructure (i.e bitcoin's) is a robustness contributor factor. Once we have established the driving variables, we need to study the volatility of those variables (and the non-linear relationships). We see how with high volatility the predictor performs poorly, at least with this small subset of variables, which are also all highly correlated. A possible fragility test would be to derive the "volatility of volatility" where the price volatility prediction becomes a traffic prediction problem. But the key realization, in alignment with Taleb and Douady, is that to be fragile the system has to be non-linear to harm (has to be accelerated to harm). Since fragility that comes from linearity is immediately visible, the hidden risks and potential harm come from non-linearity. In this example, the AI notes that the activity in a particular mining pool frequented by the users of the exchange has a low correlation to volatility, and will use the second derivative of volatility respect to the mining pool (usage per day is the event size) as a "trust" distance metric. The formula is obtained using symbolic regression as well, computing a smoothing spline for volatility with respect to mining pool, and then, computing the symbolic derivative of the spline functions (cubic polynomials), and evaluating the expression at various data points. One of the possible models has the form $D(Volatility, (Mining\_pool), 2) = (\sin(1.27652458974758 + 1.63551681837977e{-5} * (Mining\_pool)) + \cos(3.2702109619178e{-5} * (Mining\_pool) + \sin(2.14970699758616e{-5} * (Mining\_pool))))/(Mining\_pool)$

Of course, the data set can be enhanced with multiple data sources, and the prediction error reduced by combining additional methods (e.g. de-noising with recurrent neural networks and convolutional neural networks). But by providing a context of the environment to the intelligent agent, the AI is in a better position to reason on the trustability of the result.

Spatio-temporal patterns in blockchain networks

Space of Production

In one of the first studies of the disciple of Human Geography into distributed shared ledgers, Blankenship[8] conceives blockchains as production spaces where developers are the dominant class within the social and technical spaces of the technology, have ultimately leveraged their knowledge and power dynamics to accumulate wealth via the token value, and then shifted into the role of investors. This necessarily involves automation (exploitation of automated robot labor) and obfuscation of the mechanisms of production —geographic borders are defined via conflicting abstract conditions (social, political, and economic), and put within the qualitative context of social dynamics. Humans not only trust in the source, but they also trust the structure —you generally do not care about who wrote a diet article (even if a change in lifestyle can have a lasting impact on health) as long as "structure" suggest the writer is not a charlatan. A similar behavior is observed in crypto markets, where traders and investors keep lists of Twitter accounts that they trust on to relay accurate information about the state of the market, and that is facilitated by other traders: it does not count only who is saying it, but who is following —this is part of the social fabric of crypto markets, the structure of the network encodes tacit knowledge and reflects abstract conditions and boundaries. The distance trust metrics have very tangible implications for individual and corporate purposes; "a member of my group said" (even if he had materially different attributes) is generally better than what an outsider says. The implications in terms of the theory of the firm:[9] you do business in the proximity of your circle (your trust space) where trust is secured, even if it is more expensive to produce in your inner circle, and it is cheaper to acquire in the boundary (e.g. potential partners) —but going beyond that will require a significative leap of faith and the associated risk should be priced-in. The AI will understand this human inclination (as shown in Figure 2), as a trust differential in terms of metric entropy.

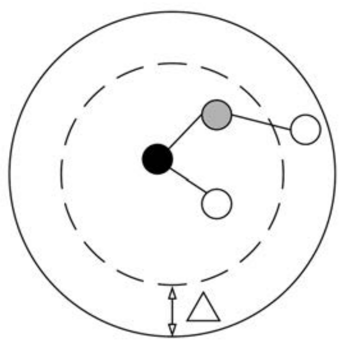


Figure 2: Trust spaces

Circles of trust and Kolmogorov entropy

Entropy is the price of structure.[10] While the models obtained using genetic programming (Figure 1) have an associated complexity which is strictly related to leaf tree structure (genotype), as seen in Figure 3, the phenotype (the expression of the genotype in mathematical formulae) itself serves as an indication of the level of complexity.

The Kolmogorov complexity measures the length of the shortest program required to reproduce a pattern. We approximate the Kolmogorov complexity using a lossless compression technique. In this way, we found that model 1 and 2 have both an approximate byte count of 416 (despite the delta in evolutionary model complexity), and model 15 has a byte count 768. That is, in (metric) information theoretical terms the first two simplest models present invariance with respect to affine transformations in the trust space. This is a disambiguation aid that the AI uses for decision making: a description of the world with a lower error (model 2) can be encoded at the same level of computational complexity as an inferior alternative.

volatility

	Complexity	$1-R^2$	Function
1	11	0.319	$10.33 + (3.02\times10^{-6})\,\text{coinbase•support}$
2	15	0.285	$1.46 + 0.24\sqrt{\text{gdax•support}}$
3	19	0.257	$3.04 + (2.44\times10^{-6})\,\text{coinbase•support} + (5.44\times10^{-6})\,\text{mining•pool}$
4	23	0.252	$-2.21 + (1.96\times10^{-6})\,\text{coinbase•support} + (7.77\times10^{-3})\sqrt{\text{price•info}}$
5	26	0.243	$-0.79 + (2.92\times10^{-6})\,\text{coinbase•support} + \frac{10.57\,\text{price•info}}{\text{coinbase•support}}$
6	30	0.234	$0.43 + (2.21\times10^{-6})\,\text{coinbase•support} + (2.89\times10^{-6})\,\text{price•info} - (8.55\times10^{-13})\,\text{mining•profitability•info}\,\text{price•info}$
7	31	0.226	$0.97 + (2.10\times10^{-6})\,\text{coinbase•support} + (2.33\times10^{-6})\,\text{price•info} - (2.91\times10^{-14})\,\text{price•info}^2$
8	34	0.221	$1.64\times10^{-13} + (2.12\times10^{-6})\,\text{coinbase•support} + (2.10\times10^{-6})\,\text{price•info} - (9.65\times10^{-21})\,\text{mining•profitability•info}\,\text{price•info}^2$
9	38	0.212	$1.74\times10^{-13} + (2.75\times10^{-6})\,\text{coinbase•support} + (1.48\times10^{-6})\,\text{price•info} - \frac{(1.82\times10^{-15})\,\text{price•info}^3}{\text{mining•pool}}$
10	42	0.207	$10.71 - (1.15\times10^{-5})\,\text{mining•pool} + (5.95\times10^{-13})\,\text{coinbase•support}\,\text{mining•pool} + (4.99\times10^{-6})\,\text{price•info} - (7.17\times10^{-14})\,\text{price•info}^2$
11	45	0.200	$2.20\times10^{-13} + (1.55\times10^{-6})\,\text{coinbase•support} + (7.27\times10^{-19})\,\text{coinbase•support}\,\text{gdax•support}\,\text{mining•pool} + (2.82\times10^{-6})\,\text{price•info} - (4.17\times10^{-14})\,\text{price•info}^2$
12	46	0.198	$2.24\times10^{-13} + (1.32\times10^{-6})\,\text{coinbase•support} + (6.02\times10^{-21})\,\text{coinbase•support}^2\,\text{mining•pool} + (3.03\times10^{-6})\,\text{price•info} - (4.66\times10^{-14})\,\text{price•info}^2$
13	48	0.193	$1.67\times10^{-13} + (2.13\times10^{-6})\,\text{coinbase•support} + (1.46\times10^{-18})\,\text{coinbase•support}\,\text{gdax•support}\,\text{mining•pool} + (1.67\times10^{-6})\,\text{price•info} - (1.77\times10^{-19})\,\text{gdax•support}\,\text{price•info}^2$
14	59	0.183	$1.86\times10^{-13} + (1.56\times10^{-6})\,\text{coinbase•support} + (1.80\times10^{-18})\,\text{coinbase•support}\,\text{gdax•support}\,\text{mining•pool} + (2.73\times10^{-6})\,\text{price•info} - (1.66\times10^{-13})\,\text{mining•pool}\,\text{price•info} - (1.80\times10^{-19})\,\text{gdax•support}\,\text{price•info}^2$
15	60	0.181	$1.79\times10^{-13} + (1.78\times10^{-6})\,\text{coinbase•support} + (1.85\times10^{-18})\,\text{coinbase•support}\,\text{gdax•support}\,\text{mining•pool} - (9.55\times10^{-20})\,\text{mining•pool}^3 + (2.33\times10^{-6})\,\text{price•info} - (2.09\times10^{-19})\,\text{gdax•support}\,\text{price•info}^2$

Figure 3: Models (BTC volatility in GDAX)

Configuring blockchain protocols' parameters based on the networks' topology analysis

Authors have explored graph-based techniques to automatically detect realistic decentralized network growth models from empirical data[11] and to study systemic risk in cryptocurrency markets.[12] The causal inference network in Figure 4 uses a parametric approach based on symbolic regression to derive the relationship between nodes, starting from a sample of 2000 price and consumption of services in the cryptocurrency markets during the period of August 2016 to January 2018.

This approach is based on empirical data, is robust to error, and has high explanatory power (all sensitivity and error-complexity figures are accessible via the description of the symbolic regression evolutions). However, to capture the full scope of causality, the "inherited fragility", the AI makes use of signal processing and other techniques. Figure 5 depicts the use of the wavelet coherence method, which has been previously applied to the study commodities and financial time series[13,14] to understand when and how strongly an off-chain signal (in this case, the usage of a popular Ethereum web wallet service) affects the price of the cryptocurrency, ether. The AI uses this not only for disambiguation but to actually map the causal relationship in time and frequency domain (i.e. learn when one signal lags or leads the other).

Here the wallet signal leads the price signal, on day 120 in a cycle of approximately 4–6 days, where both signals are also strongly correlated. In the case of unruly distributions, the results are enhanced in combination with other methods (e.g. Granger causality,[15] Bayesian structural time series models,[16] among others). The key realization is that although market behavior (network dynamics) is not the same as market conditions (market structure), the intelligent agent can use the same graphical metaphor to gain context, and predict the counterfactual response. But the risk is that correlation at training time does not ensure correlation at test time, the AI should therefore be aware of divergences.

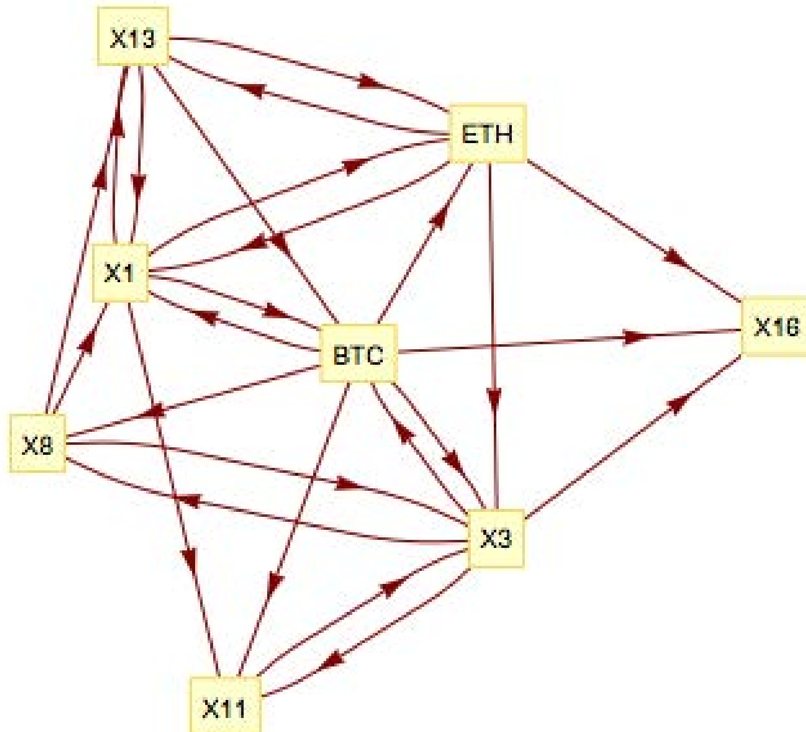


Figure 4: Causal inference network (BTC-ETH)

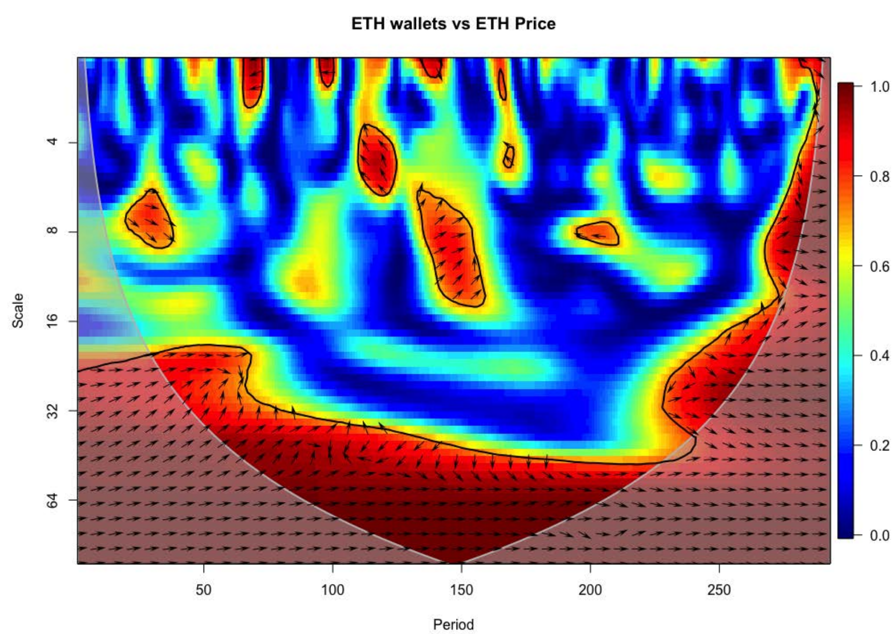


Figure 5: ETH wallet vs ETH price. July 7th 2016 to June 30th 2017.

Use of low complexity property testing methods by decentralized blockchain agents

Neighborhood Asymmetry

Once the AI is context aware (of the trust space in terms of entities, and of the relevant variables given various causality tests) information metrics derived from the data space itself are utilized to measure the actual information content of each sample. The neighborhood asymmetry method[17] sums the vectors from the data record to the neighbors implicitly defined by the supplied data matrix and returns the length of this resulting neighborhood directionality normalized by the number of neighbors. In this way, this metric is primarily concerned with the symmetry of the neighbor distribution but also contains a contribution from the distance to each of the neighbors. Figure 6 shows the symmetry for the BTC off-chain economy modeled in terms of on-chain economy variables (fees) in the period of August 2nd, 2017 to January 24th, 2018; it tracks over-the-counter exchanges (OTC), wallet services, paper wallet generators, block explorers, among many others.

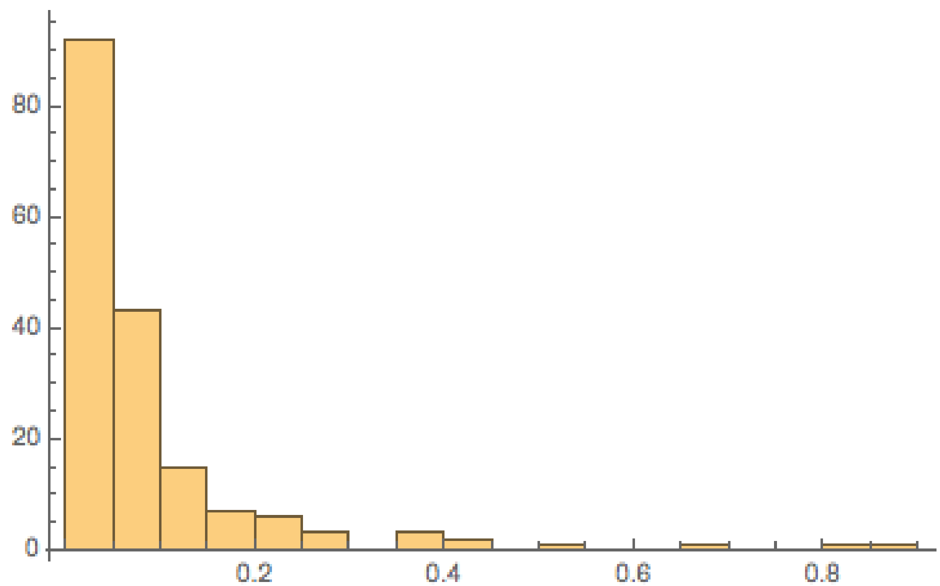


Figure 6: Histogram (BCH)

We are concerned with measuring the maximum distance (infinity-norm) between any one of the variables. We see that most data points have larger metrics, with a few points breaking the symmetry (e.g. 0.8–1 bin). Since we are interested in material connections between the asymmetry of the neighborhood data space and actual trust asymmetry, we use fees as response variable (an actual on-chain transaction metric), alongside the behavioral signals of off-chain economic and investment activity.

Mathematical invariants

Ensembles of models of diverse but comparable performance and complexity lead to a trust metric.[17] Figure 7 shows how the intelligent agent *perceives* the trustability of the prediction, what may happen in regions of unknown parameter space (when it is exposed to unseen data) or when the underlying system changes. The AI naturally finds interesting those inputs that show invariance, as well as the points where the symmetry breaks for the others.

Ensembles are constrained to diverge. The trustability of the prediction (i.e. an assessment of the confidence in the prediction) is measured using an ensemble divergence function, which captures the response consensus behavior of the supplied model ensemble. Figure 8 shows all possible combinations of variables displayed as a 3D surface, with the spread in the embedded models reported as 3 standard deviations.

Behavioral finance

Mis-pricing due to non-rational decision making and market inefficiencies offer arbitrage opportunities that an intelligent agent would try to exploit. But an AI that is aware of the trust asymmetries across its operational space is also able to compare intrinsic fragilities, even if the distribution of problematic events are not directly observable. In Taleb and Douady

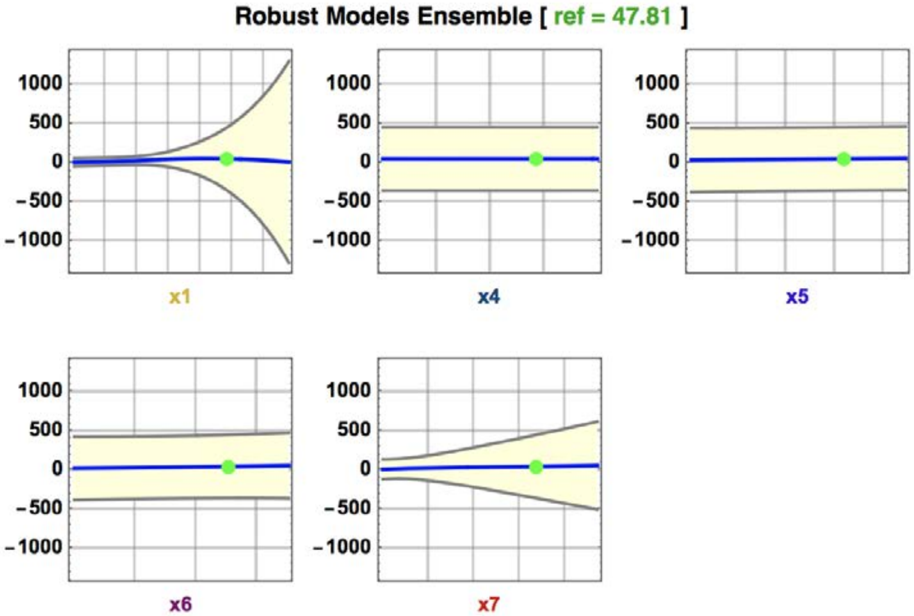


Figure 7: Ensemble response plot (BCH fees)

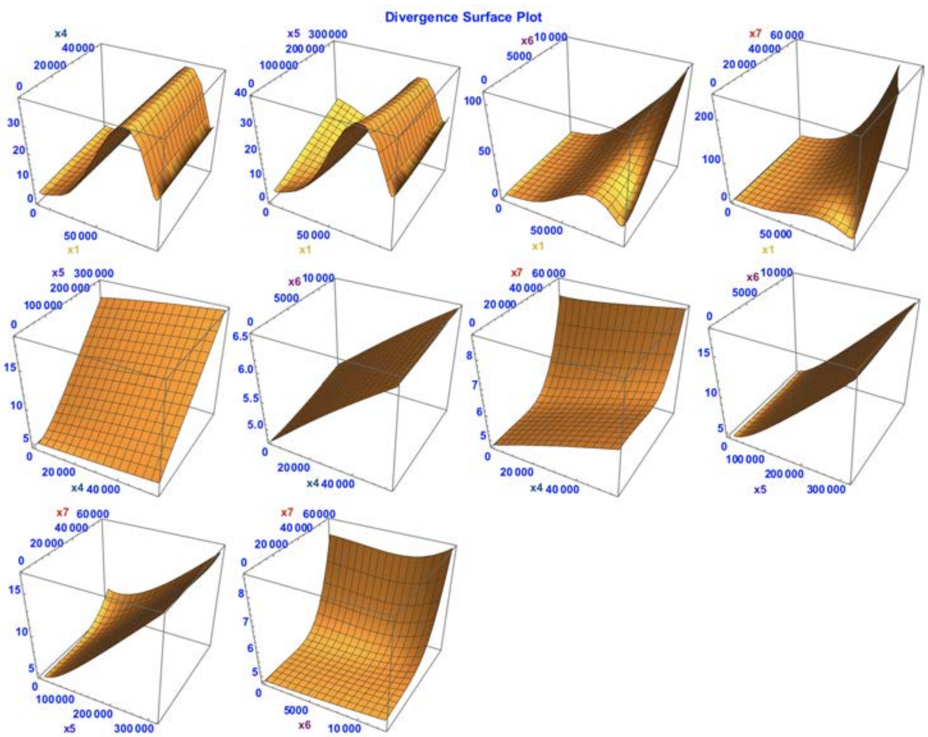


Figure 8: Divergence (BCH)

formulation, the risk measure assumes certain extrapolation rules that have first-order consequences, and these consequences are even more amplified when the risk measure applies to a variable that is derived from the one used for estimation —when the relation between the two variables is strongly nonlinear. For (predictive) navigational purposes, the agent asks whether people (or other AIs) have found a path of least resistance (evidence of trust asymmetry) and what is its associated trustability. This aspect is modeled using Forrester dynamics, and it captures the response to changes satisfying both system modeling and risk modeling.

Information flow

The AI that seeks to optimize blockchain configurations, or simply navigate the environment, is aware of the off-chain "social fabric" because it fills the role of communication: you can improve by introspection if you communicate.

To manage complexity several approaches are possible, including using Lawyer's expected force (a centrality measure)[18] to greatly simplify the problem. When node power is low, influence is a function of neighbor degree. As power increases, a node's own degree becomes more important. The strength of this relationship is modulated by network structure, so it is expected that it will be more pronounced in narrow and dense networks such as social networking (e.g. Twitter channels of bitcoin whales). The network effect, however, has two levels: one is the on-chain\off-chain interplay (symbolic regression model of active accounts/addresses driven by social network activity), and the other is the off-chain\off-chain interplay (for instance, of the dynamics between Twitter channels and Telegram groups).

Vector fields as temporal streams

An activity must be decoded sequentially over time. The intelligent agent may do this by using a combination of genetic programming techniques (after all, somewhat static DNA and its transcription pattern over time creates biologically essential temporal patterns) and signal processing (e.g. Kalman filter, for short-term streams), but a complementary approach for fast evolving systems that are always in flux is to use actual fields. In the case of economic systems such as blockchains, standard signal processing analysis techniques and information theoretical measurements help visualize the historic correlations, but a sound investment strategy should also consider the correlation migration —how the correlation changes (or not) over time. While it is possible to plot a correlation graph for each point in time, we find that using vector fields[19] allows for mappings with a higher information density, especially for portfolios of a large size —for instance, consider that when tokenized Dapps are also viewed as assets, we are facing prohibitory large portfolio sizes. To implement the method we begin by defining the convention for the vector components. From the possible traffic sources for a new project, we found that referrals and social networks are the more prevalent, especially in the early stages of a proposal listing when word of mouth in social networks such as Reddit and the ability of the founding team to generate buzz in media and news sites plays a role. The resulting vector field gives rise to a flow. A fluid flow provides an effective way to summarize the dynamics of a portfolio to include an arbitrarily large number of entities, rather than simply scaling up the number and size of correlation graphs (not to mention that for communication purposes, a vector field is also a more intuitive representation of cashflows equivalents). In one hand, the (total) vector magnitudes are a measure of strength, in the other, the interaction between the different assets (as revealed by singularities in the flow) present a portrait of the system (the portfolio attention correlations). Figure 9 shows a vector field for a 32-asset mapping in Ethereum during March, April, and May 2016, rendered using 4 techniques to highlight different aspects of the flow.

Investors (and the intelligent agent) ascribe an intrinsic value to stability. By using the mapping from Figure 9 and small multiples to draw each month separately, every field snapshot is a moment in time, and the apparent flow mobility shows progression in portfolio positions. Therefore, one can easily identify which flow structures tend to remain unchanged, and when a major event occurred; the streamlined plot (Figure 9.d) is ideal for such type of visual analysis. The vector plot (Figure 9.c) presents the "intensity" dimension lacking in the streams, which are focused on directionality. The LIC (line integral convolution) rendering (Figure 9.b) is a human friendly and aesthetically pleasant format that helps reinforce the flow structure without losing analytical capability, especially if one makes it overlap the vector map, or the streamlines, depending on the data dimension to analyze (by using LIC the entropy of the visualization increases, more information is conveyed). The mesh network representation of the flow (Figure 9.a) is a machine format that maintains some of the human friendly features of the other visualizations —one can easily see how computation by clusters might become a useful device to reveal equilibrium points in the vector field topology: unstable nodes (sources or saddles), stable focus (spiral sink), stable centers, etc. This representation is obtained by drawing mesh divisions between every line or polygon generated by a plotting function, in this case, the one obtained after the stream plot. Finally, given the duality between singularities in vector fields and network structure, the fields analysis is suitable to be implemented at scale.

Applications

Intelligence services

Although privacy coins offer desirable features in some settings, for national security purposes is often needed to understand the user's clusters at least at a macro level. Trust asymmetry (in the blockchain / off-chain boundaries) reveals information

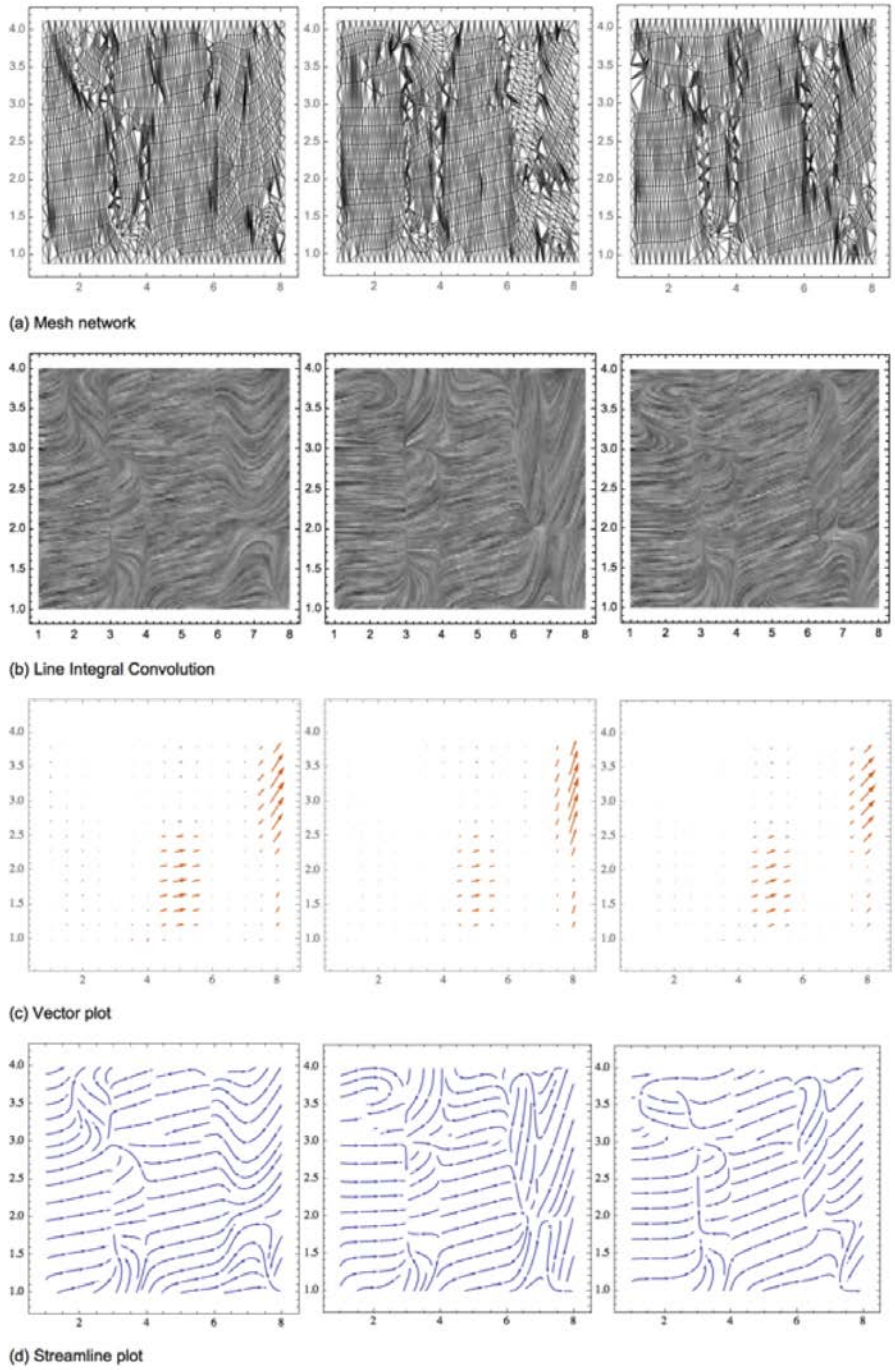


Figure 9: Time variant vector field sequence

such as geography and audiences, which allows to re-construct digital personae (groups' identity profiles); it also provides insights into hidden states and phase transitions within those multiplex networks. The applications include mapping the character of Zcash and Monero communities and providing partial source/destination metadata related to the use of secret contracts and similar technologies that facilitate coin mixing (which obscure the original source of cryptocurrency used within the protocol). These applications are also relevant to market intelligence, a practice that becomes more challenging as international privacy standards strengthen.

Financial Risk prediction

Ensemble prediction with evolutionary computing augments both technical analysis and time series ARNN results by providing context and by providing explicit trustability ranges. The intelligent agent is thus capable of reasoning about interventions: what if I change X? This is the alternative to creating static models —rather, we create models dynamically and perform spatiotemporal encoding of information using a blockchain as a substrate for the intelligence. This also implies a sort of "value by memory" where a system that has *experienced* more about the world is more valuable than one that consists only of the algorithm.

Political risk prediction

The political economy is a contest for attention modulated by reputation, and thus a perfect fit for context-aware AI.

In multiples areas, from influencing public opinion to political campaign financing, demand is influenced by the available supply (of information).

An AI that minimizes trust asymmetries on its objective function can use well-known human biases (tendency to follow rules of thumb-credulity, cognitive dissonance-double down on beliefs, human inclination to spread pleasant-sounding lies) to assess provenance of the data and context, in applications related to detection and mitigation of fake news and deep fakes. The embedded Oracle can operate as a standalone feed, or as a complement to specialized technologies that take advantage of blockchain features for this purpose (no single arbiter of truth, public record). In this application, the genetic programming provides an audit trail on its genome (tree structures), and, an event-based notification function (when a breakdown of trust symmetry occurs).

Part of the tasks of the AI will deal with inferring intent: for instance, an analysis of the Bernie Sanders and Gary Johnson campaigns shows asymmetries at higher scales: the demand for information and actual voter commitment do not correlate for the top contributors (attention price). But It also pertains to crypto native applications: the bitcoin donation adoption was found to be more prevalent among libertarians, a group ideologically aligned to decentralization of monetary policy.

Conclusions

The characteristics of blockchain-based AIs such as being incentivized natively through the use of tokens of value, and not having a single point of failure, are attractive propositions, but also mean that decentralized intelligence will be hard to kill if something goes wrong. And, depending on the stage of its development, such a decision could meet ethic questioning. It is therefore imperative that those intelligent agents go beyond the basic expectation (do not do harm to humans, do the job, do not lose money) to actually solve the vulnerability issues of humans systems (security) while easing human anxieties by providing transparency (in the words of Manuela Veloso, verifiable answers and consistency of answer) and operating under a set of beliefs ("mental" models) that are compatible with the human experience. To do this, the AI needs to have an adequate degree of trustworthiness on its own assessment, and trustable symbolic regression provides a means to that —in the same way that humans augment their intelligence with AIs, AIs can augment their intelligence with a time-variant model of the environment created from off-chain signals. This implies both operating with a reasonable degree of intuition about cause and effect, and with the ability to deal with edge cases (even in the case of narrow, purpose-specific AI).

If we take a lesson from history, making a new weapon so terrifying that it will be inconceivable to use (e.g. Leonardo's battlefield tanks, von Neumann's mutual assured destruction nuclear doctrine), is itself a form of deterrence. If blockchain is truly irreversible social computing,[20] and being trust hard to earn and rebuilt, memory itself could be a useful deterrent for misbehavior and carelessness, for humans and machines alike. But it also means that where there is a trust imbalance (usually in the periphery of the blockchain, in the coupling with the off-chain systems that support it) there are opportunities for either trust disintermediation or arbitrage, and possibly, value creation. Moving forward, this combination of awareness of irreversibility, value by memory, and reasoning about introspection, perhaps implemented using non-ergodic variants of cultural genetic algorithms,[21] could allow machines to navigate the world using the same fundamental device that evolution has provided to humans: trust. And ultimately, the question is not if we should trust AIs, but rather how AIs will trust us.

References

1. Kyburg, H. E. & Pearl, J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. *The Journal of Philosophy* **88**, 434 (1991). URL `https://doi.org/10.2307/2026705`.
2. Wölfer, R., Cortina, K. S. & Baumert, J. Embeddedness and empathy: How the social network shapes adolescents' social understanding. *Journal of Adolescence* **35**, 1295–1305 (2012). URL `https://doi.org/10.1016/j.adolescence.2012.04.015`.
3. Liu, Y.-Y., Nacher, J. C., Ochiai, T., Martino, M. & Altshuler, Y. Prospect Theory for Online Financial Trading. *PLoS ONE* **9**, e109458 (2014). URL `https://doi.org/10.1371/journal.pone.0109458`.
4. Engelbart, D. C. Toward augmenting the human intellect and boosting our collective IQ. *Communications of the ACM* **38**, 30–32 (1995). URL `https://doi.org/10.1145/208344.208352`.
5. Fareri, D., Chang, L. & Delgado, M. Computational substrates of social value in interpersonal collaboration. *Journal of Neuroscience* **35**, 8170–8180 (2015). URL `https://doi.org/10.1523/JNEUROSCI.4775-14.2015`.
6. Venegas, P. Trust Asymmetry. URL `https://doi.org/10.22541/au.151979448.82088697`.
7. Taleb, N. N. & Douady, R. Mathematical definition mapping, and detection of (anti)fragility. *Quantitative Finance* **13**, 1677–1689 (2013). URL `https://doi.org/10.1080/14697688.2013.800219`.
8. Blankenship, J. *Forging Blockchains: Spatial Production and Political Economy of Decentralized Cryptocurrency Code Spaces*. Master's thesis, University of South Florida (2017).
9. Foreword: Coase and the Theory of the Firm. *Managerial and Decision Economics* **36**, 1–1 (2014). URL `https://doi.org/10.1002/mde.2700`.
10. Prigogine, I., Stengers, I. & Pagels, H. R. Order out of Chaos. *Physics Today* **38**, 97–99 (1985). URL `https://doi.org/10.1063/1.2813716`.
11. Menezes, T. & Roth, C. Symbolic regression of generative network models. *Scientific Reports* **4**, (2014). URL `https://doi.org/10.1038/srep06284`.
12. Venegas, P. On the sources of systemic risk in cryptocurrency markets. URL `https://doi.org/10.22541/au.152074266.61910113`.
13. el Alaoui, A. O., Dewandaru, G., Rosly, S. A. & Masih, M. Linkages and co-movement between international stock market returns: Case of Dow Jones Islamic Dubai Financial Market index. *Journal of International Financial Markets, Institutions and Money* **36**, 53–70 (2015). URL `https://doi.org/10.1016/j.intfin.2014.12.004`.
14. Nagayev, R., Disli, M., Inghelbrecht, K. & Ng, A. On the dynamic links between commodities and Islamic equity. *Energy Economics* **58**, 125–140 (2016). URL `https://doi.org/10.1016/j.eneco.2016.06.011`.
15. Auinger, F. VIX Index. In *The Causal Relationship between the S&P 500 and the VIX Index*, 37–52 (Springer Fachmedien Wiesbaden, 2015). URL `https://doi.org/10.1007/978-3-658-08969-6_6`.
16. Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N. & Scott, S. L. Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics* **9**, 247–274 (2015). URL `https://doi.org/10.1214/14-aoas788`.
17. Vladislavleva, E., Smits, G. & Kotanchek, M. Computational Intelligence in Industrial Applications. In *Springer Handbook of Computational Intelligence*, 1143–1157 (Springer Berlin Heidelberg, 2015). URL `https://doi.org/10.1007/978-3-662-43505-2_57`.
18. Lawyer, G. Understanding the influence of all nodes in a network. *Scientific Reports* **5** (2015). URL `https://doi.org/10.1038/srep08665`.
19. Čižinská, R., Krabec, T. & Venegas, P. FieldsRank: The Network Value of the Firm. *International Advances in Economic Research* **22**, 461–463 (2016). URL `https://doi.org/10.1007/s11294-016-9604-x`.
20. Rao, V. Blockchains Never Forget. https://www.ribbonfarm.com/2017/05/25/blockchains-never-forget/. URL `https://www.ribbonfarm.com/2017/05/25/blockchains-never-forget/`. Accessed on Mon, May 28, 2018.
21. Reynolds, R. & Liu, D. Multi-objective cultural algorithms. In *2011 IEEE Congress of Evolutionary Computation (CEC)* (IEEE, 2011). URL `https://doi.org/10.1109/cec.2011.5949757`.

Portuguese Venture Capital Ecosystem - Visualising actors and investment flows across time

Diana Vieira Fernandes[1*,2]

[1]IN+, Center for Innovation, Technology and Policy Research,
Laboratory of Robotics and Engineering Systems (LARSyS),
Instituto Superior Técnico, Universidade de Lisboa,
Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
[2]Nova Law School, Universidade Nova de Lisboa, Lisboa, Portugal

*diana.vieira.fernandes@tecnico.ulisboa.pt

The aim is to present the Portuguese Venture Capital Ecosystem, over time, using the national ecosystem data to understand the investment type under analysis. Venture Capital is characterised by investment in (new) companies ("start-ups"), unlike other investments, including in asset classes with greater liquidity, Venture Capital, does not have as characteristic, returns with a normal distribution. Similarly, when designing cities, houses or infrastructure one does not use the weather data from places with considerable different weather and population patterns (i.e. north pole data to build a city in southeast Asia). Still, in this case, we keep assuming that the empirical evidence of different systems (and categories) will fit other jurisdictions and markets. Many studies carried on this type of investment in Portugal, usually choose to analyse a given particular case or, use the data of all aggregated investment in Venture Capital and Private Equity. Secondly, most perform analysis over annual reports and not over an extended period. Network analysis may be used as a tool for a better understanding (and contract modelling) of Venture Capital Investment.

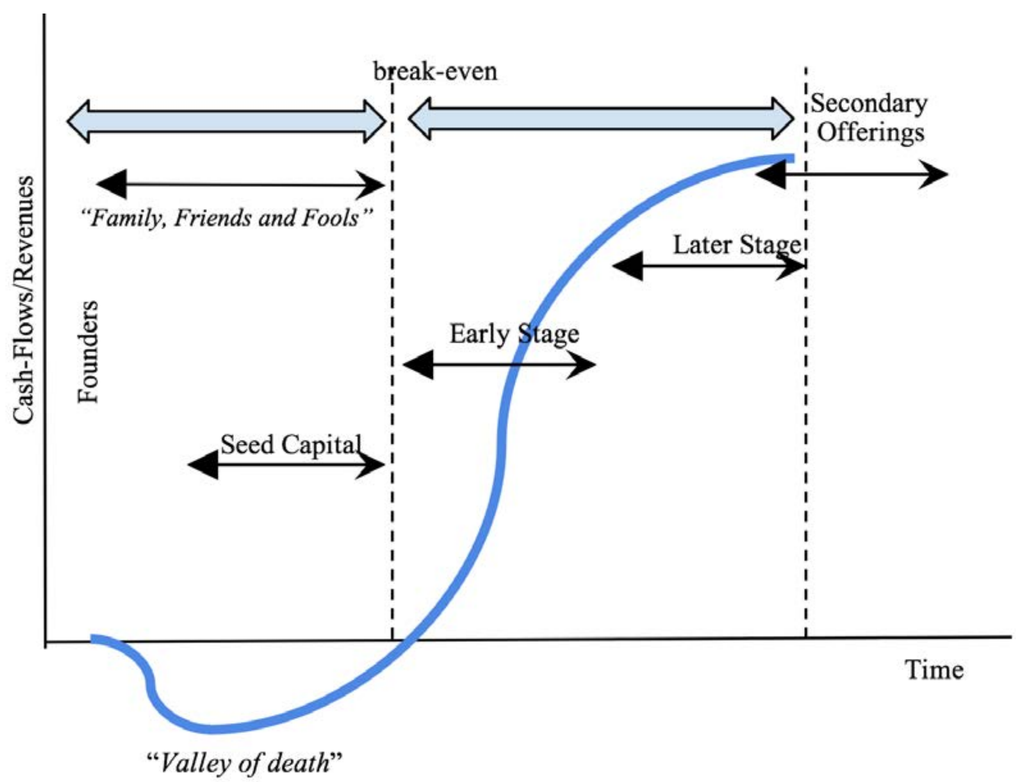
1 Introduction

1.1 Venture Capital definition

Venture Capital is characterised by investment in companies (”start-ups”, where this category can be subdivided into ”seed capital”, ”start-ups” or ”Early stages” and ”later stages”, usually with less than 3 to 5 years of existence, in the case of ”seed capital”), particularly with a strong technological component. As such, has as common traits, accumulated losses in the early years, which may reflect investment in R&D. Before a new product or service is introduced in the market, goes through several stages of development and validation, where the ”Technology Readiness Levels (TRL)” [1] is usually considered as the evaluation scale of the product development phase.

As shown in Figure 1 below, the venture capital financing operations [2], target companies in the early stages, in particular, when they are not yet able to generate revenues.

Figure 1: Start-up Financing Cycle



A large part of the products and services that these new companies develop, namely, if they are in an embryonic stage of development (in addition to the specificities of specific areas, in which the pharmaceutical sector assumes a greater expression, related to specific procedures for research and commercialisation) do not generate any significant income in the short term. Thus, incurring in cumulative and sometimes substantial losses (i.e. high costs in research, IP rights and registration processes, so as in all phases before commercialisation sometimes referred to as "proof of concept" or, in particular cases mentioned above, in clinical trials) before generating any revenues.

It should be noted that although there is a potential for significant returns[1] - if the technology passes through the various stages of validation [2] - an investor, the more the product is in an embryonic state, the greater the uncertainty [3], namely technological. For this reason, the investment is also made in stages, usually called "investment rounds" to reduce and mitigate investment and technology risks. In addition to the inherent uncertainties, due to the characteristics of the product/service itself, there are also the uncertainties regarding the product's acceptance. Namely in products based on the need for a large number of users, such as digital platforms, where the network economies are a vital part of the success (or failure) of the product/service and, finally, the external dynamics, namely competition from other companies and regulatory issues in specific markets.

Unlike other investments, including in asset classes with greater liquidity (as the case used by Harry Markowitz for the formulation [4] of Portfolio Theory [5]), Venture Capital, does not have as characteristic, returns with a normal distribution [5], besides the funds having (usually) a 10 years' term, implying that they must disinvest (exit) prior this date. Besides the diversification methods - also imposed by the CMVM [6] - Regulation or, even through the syndication of the investment, in order to mitigate the risk, it is considered that a given investment with a rate of return lower than other types of investment with lower risk (where at the other extreme are the allocation of capital to bank deposits or investment in treasury bonds) are considered unsatisfactory. On the other hand, since a large part of the capital investment is made by the investors (percentage), the loss compared to the founding partners will, in most cases, be exponentially greater. This notion is of the greatest importance, not only on financial modelling, but also, on legal grounds[7]. When formulating rules (which may emanate from regulations, laws, contracts,

[1] I.e. according to Capital Asset Price Model (CAPM), beta is the only relevant measure of a stock's risk. It measures a stock's relative volatility (i.e. shows how much the price of a particular stock moves up and down compared with how much the stock market, as a whole, moves). Black, Fischer et. al. [3] studied the price movements of the stocks on the New York Stock Exchange between 1931-1965 and confirmed a linear relationship between the financial returns of stock portfolios and their betas.

[2] In the literature, it is also called "R&D pipeline", that is, the initial number of ideas, concepts that are developed, many are being discarded for several reasons, one of them being the fact that a certain (technical) hypothesis is not validated. The number of ideas that are materialised and have a financial return is thus reduced.

[3] On the difference between risk and uncertainty, Cf. Knight, F. H. Risk, Uncertainty, and Profit. (1921) [4]

[4] Where it uses the variance of asset prices as a proxy for risk (also, known as mean-variance analysis).

[5] Most use pricing models that imply an assumption on revenues (the concept of expected return), wherein this case due to the lack of historical data, would be treated as an uncertainty (event not known/occurred) and not as risk (known event). Or the distinction between value and price. [6]

[6] Comissão de Mercado de Valores Mobiliários (CMVM) or, Portuguese Securities Market Commission.

[7] Cf. "Venture Capital Method" by Sahlman, William A., and Daniel R Scherlis. A Method For Valuing

which are abstract by nature), are rarely analysed the impacts and effects of a certain rule, in extreme events[8].

In the contract design [7] [8], there are different general solutions to the "abnormal" cases. The regime of the change of circumstances, the definition of contracts as random or, lastly, as "incomplete contracts" [9]. In the case under analysis, the uncertainty (event) is known[9], by the parties, although, mode and amplitude (measurable risk) and, the time of the occurrence are unknown [4] [9].

However, taking these cases, the remaining cases seem to refer to cases in which a given effect or behaviour, in a given population, has a normal distribution and does not have significant dispersion phenomena (as in the example of distributions that present observations in the tails). Nor of readjustment (change in the weighting of explanatory or predictive variables), in fact, it is the average as the standard deviation, which is often used to limit the behaviour of acceptable solutions[10]. However, venture capital investment does not comply with such a rule, nor does it intend to invest in companies that have "normal"[11] returns. The same can be said of the so-called "successful companies"[12], which are more reported[13], than the unsuccessful ones.
According to the INE[14] (Statistics Portugal) methodological notes[15], are:

- **High-growth enterprises**: Companies (with 10 or more paid employees) with an average

High-Risk, Long-Term Investments: The "Venture Capital Method". Harvard Business School Background Note 288-006, July 1987. (Revised October 2009)

[8]It ends up having a similar behaviour (distribution) to cases of risk analysis in insurance (both diversify, to mitigate risk and exposure). The only difference in risk analysis is that is intended to avoid this excessive exposure to an event with a potential negative impact (i.e., only insuring land in an area with heavy rainfall and the occurrence of a flood), there must be resources to cope with these events. In venture capital, looking to this extreme event (companies that have abnormal multiples of return), to compensate the losses of the (under-performing) ones of the portfolio (or fund). The exceptionality is in the quantity (range), not in the phenomenon (existence).

[9]The question is whether it is negligible or not. As an example, a flood is not an unknown event; it is a rare event. It has the same behaviour: when rainwater discharge meets certain conditions.

[10]On this subject check, Central Limit Theory.

[11]The average population of companies in Portugal is significantly different from those that are the subject of the type of operations under analysis.

[12]The idea is explained in "The limits to growth" (Meadows et al., 1972) [10], also transported for the analysis of the GDP growth rate of a given country. It is applicable for two reasons: if companies survive by the profits generated, it would have to be assumed that demand would be sustainable and with equal growth rates, but infinitely (i.e., there would be no deceleration). On the other hand, this demand would not be limited to the existing market. The value of companies like Alphabet (GOOGL:US), Facebook (FB:US) and others are also due to be able to capture a substantial market share. Whereas there is no market for two companies with an equal market share (it passes, the 100% of the total of the demand, generating revenue), there will be no "room" for two companies with equal value (at this point value is a function profits to simplify).

[13]Cf. contract signalling and related literature that analyses the production of exterior signals when seeking investment from third parties.

[14]Instituto Nacional de Estatística ("Statics Portugal"), url: `https://www.ine.pt/`

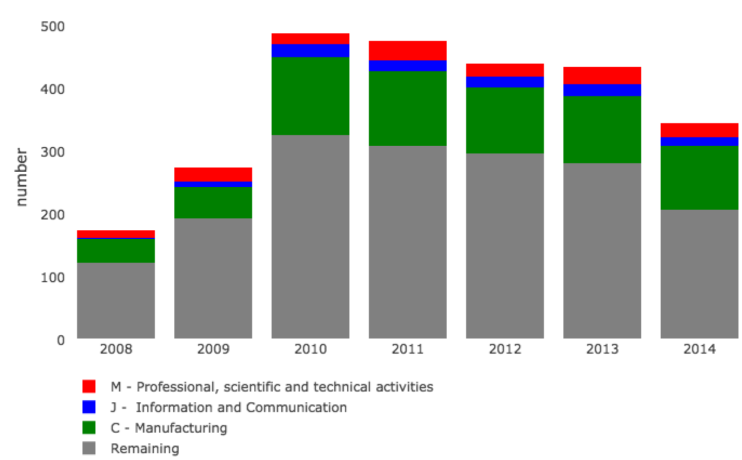
[15]INE, Business demography indicators of non-financial enterprises in Portugal

annual growth of more than 10% over a three-year period, growth is measured in terms of the number of paid employees;

- **High growth young enterprises**: High growth young enterprises - Companies (with 10 or more paid people) up to 5 years of age with an average annual growth of more than 10% over a three-year period, growth is measured in terms of the number of paid employees (Gazelles).

Using the year of 2014 as an example (and last available), in relative terms, of a total of 362 415 companies (incorporated and active), only 39 648 had more than 10 persons employed. Whereas in both cases of high-growth companies, it has a requirement of having more than 10 persons employed, of the 39 648 existing only 3415 meet the cumulatively with the requirement "with an average annual growth exceeding 10% over a period of 3 years". If added the requirement "up to 5 years of age", only 343 met the requirements to be considered "high growth young companies". Percentage-wise, less than 1% of the companies are a "high growth company" and less than 0.1% meet the requirements of "Young High growth company." When analysing[16] what are "start-ups", this classification proves to be important, to demonstrate that in relative terms we are analysing a very small part of the population.

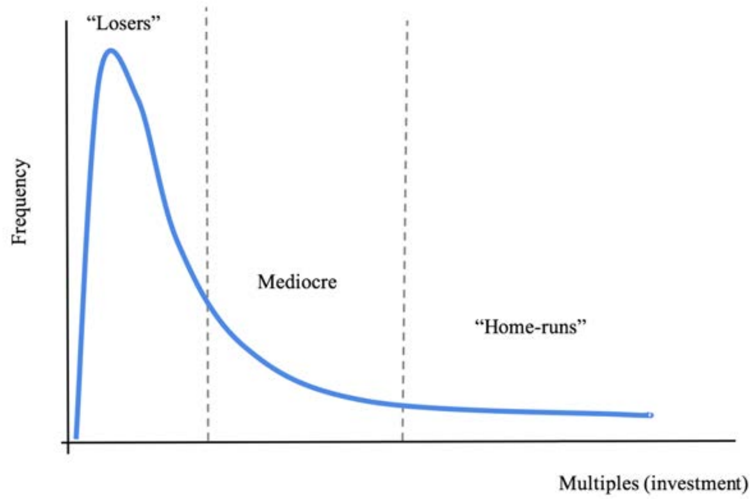
Figure 2: Gazelles" (young high growth companies) per NACE



Illustrated, in Figure 3, the y-axis represents the frequency (number of invested companies) and the x-axis the return multiples of the initial capital invested. This distribution curve is highly skewed, positive asymmetric, forming a tail on the right side, i.e., where are the highest values of the x-axis[17]

[16]This data contrast with the number of companies that had venture capital financing. On the other hand, 2011, due to the availability of funds through public risk-sharing instruments, was the year in which more venture capital

Figure 3: Successful “start-ups” Probability Distribution



Another implication of this idea is the portfolio’s valuation[18], considering the different methodologies, in the case of portfolios composed of investment units, or other instruments that have as their underlying assets, investment in ”start-ups” (either using debt, equity or hybrid instruments) namely without sales (or at least break even[19]), in an illiquid market[20] as venture capital, many of the models fail, as is later visible in the annual reports or the subsequently reported capital losses.

1.2 What is different between corporate finance and entrepreneurial finance

- Nonexistence of historical data, i.e., any projection of Cash Flows, is an estimation with no real baseline;
- Agency problems;
- Need for disinvestment and illiquidity of the market;

funds were set up.

[17] It is important to remember that in addition to the normal distribution, there are others, of which the best known are the Pareto, initially used to describe the distribution of income. The Generalised Extreme Value (GEV), which have as peculiarity the existence of extreme values (and therefore the use of non-parametric statistics), in comparison with the ones that impose linearity (if a transformation is not suitable), since most assume a given distribution (usually the Normal).

[18] Cf. Damodaran, Aswath,”Price and Value: Discerning the difference,” [6]

[19] Cf. Damodaran, Aswath, “Valuing Young, Start-Up and Growth Companies: Estimation Issues and Valuation Challenges SSRN Electronic Journal (2009)[11]

[20] Cf. Lerner, J. and Schoar, A., ”The Illiquidity Puzzle: Theory and Evidence from Private Equity” [12]

- Distribution, i.e., most companies (observations) are not successful;
- Asymmetry of information;
- Portfolios of VC Funds only hold risky assets with higher Standard Deviation than other assets.

1.3 System Characterisation

Most of the literature is based on a system (financial and legal), different from the Portuguese case[21]. On the other hand, these two ”systems” are related to each other. While the Law[22] tries to set rules and limits (and impose them as a last resort by the application of sanctions[23]), economic agents seek to maximise profit[24], depending on incentives they have.[25]

The agents, whether from a legal or economic perspective, are the same, in which they organise and function in a given ecosystem, with prescribed rules[26] and others that emerge. In the field of economic analysis of Law, it is often studied the impact that a particular rule will have a given universe of receptors (the most frequent case is the fiscal changes). The greatest difference is the recognition of the system dynamics and of the inherent entropy, which cannot be understood only through the decomposition of the system into its constituent parts.[27]

1.4 Determination of agents and interdependence between them and the system

In the venture capital ecosystem, there are two sets of major players: investors and promoters and, financiers and financial intermediaries. The way they relate is formalised through contracts, which their limits are influenced by the ecosystem, both by prescriptive (law) and emerging (financial market) rules.

[21] The most observed cases are from the USA and in the European market, the UK.

[22] The role and function of the Law will not be addressed, although it should be mentioned the two main “families” of legal systems: common law (e.g. US and UK) and civil law (as the case of Portugal and Germany), besides the Roman Law influence and the Iūs Gentium.

[23] Out of this work the role, function and end of Law, where sanctioning should be understood within each legal system, in a broader sense (not only just punitive).

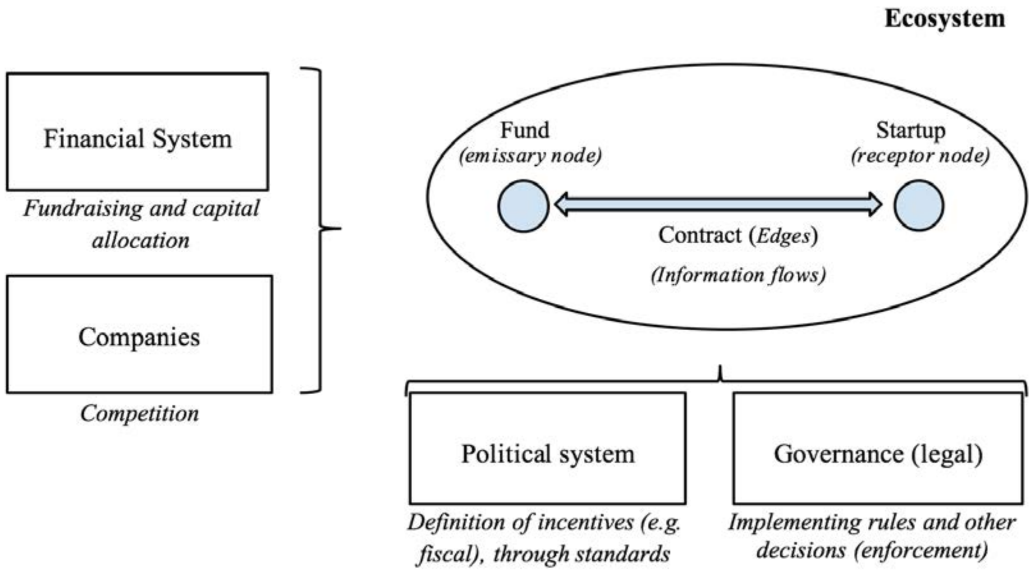
[24] e.g. “Edgeworth’s Box” [13]

[25] Cf. Opportunity Cost and Incentives, e.g. “The power of Incentives: How Seat Belts Kill, “The Armchair Economist” by Steven E. Landsburg”

[26] For example, with Acts, Laws (or other types of legislative acts), a set of norms (as Codes), namely if imperative norms.

[27] However, this exercise was done (especially after the 2008 sub-prime crisis), particularly in the risk analysis departments (including compliance), the supervisory authorities (where they look for the ”systemic risk”) and antitrust (in which they seek for excessive market concentration, or ”clusters”).

Figure 4: VC Deals Ecosystem



Contracts contain several provisions that aim to align incentives[28] and which can be applied in the Portuguese system and, taking into account the objective and possible scenarios, the two principal instruments[29] of structuring these operations are debt or equity instruments (which are limited either by law, on the one hand, and by the decisions of agents, on the other).

1.5 Money flows in Venture Capital in the system

According to Bob Zider [14], the Venture Capital industry has essentially four players: Entrepreneurs who need funds; investors who want higher returns; the Investment Banks which wants companies to sell, the capital markets to underwrite these securities and Venture Capitalists that create the market for the other three. As it will be illustrated in Figure 5, one can consider the transformation to the Portuguese case. Where without a capital market and a weak private placement market (i.e., to whom to sell these interests), to liquidate (buy side) positions, there is an additional challenge in disinvesting.

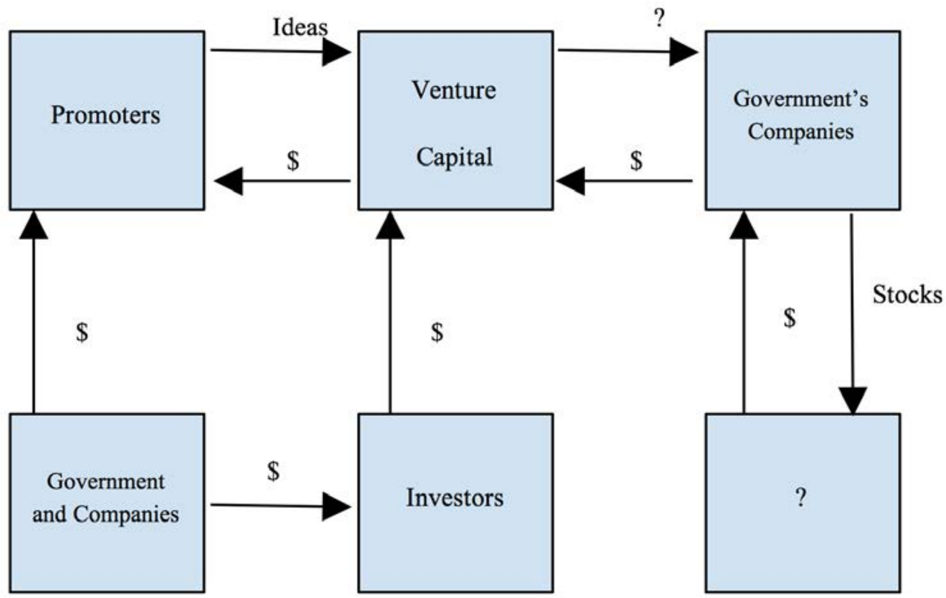
The question becomes: considering the above scheme does it makes sense to use the same model (US)[30], when one of the fundamental parts is not present? Is it wise to assume and use

[28]Incentives besides the economic definition, under a juridical interpretation, also encompasses a given expectation (fidūcia) on a future event or outcome (promises and bargains), with a certain degree of risk (business risk) the parties agreed.

[29]Also, known as "term sheets", or the set of contracts (and juridical acts) that set the terms and conditions between the parties.

[30]It is usually referred to the paper of Kaplan and Strömberg ("Financial Contracting Theory Meets the Real

Figure 5: Redesign adjusted to the Portuguese ecosystem



models in which exit strategies are formulated based on the options of selling to third parties or IPO[31]? If these scenarios are likely to be reduced[32], will it be the latter sufficient to compensate for a weak and feeble development of the capital markets?

It is important to characterise the ecosystem as well as the actors and how they interact with one another, namely the actors (agents), as the contracts (edges) between these two over time[33]. In the present work, the investments were mapped, per Venture Capital Fund, as shown in the figure 6 infra.

Whereas the system under analysis, the theory of complex systems has as its objective the connections among several objects, usually dynamically. As explained by Arthur [16] [17] corresponds to the intermediate layer between the macro economy, i.e., the aggregate behaviour of a given economy and the casuistic analysis [16].

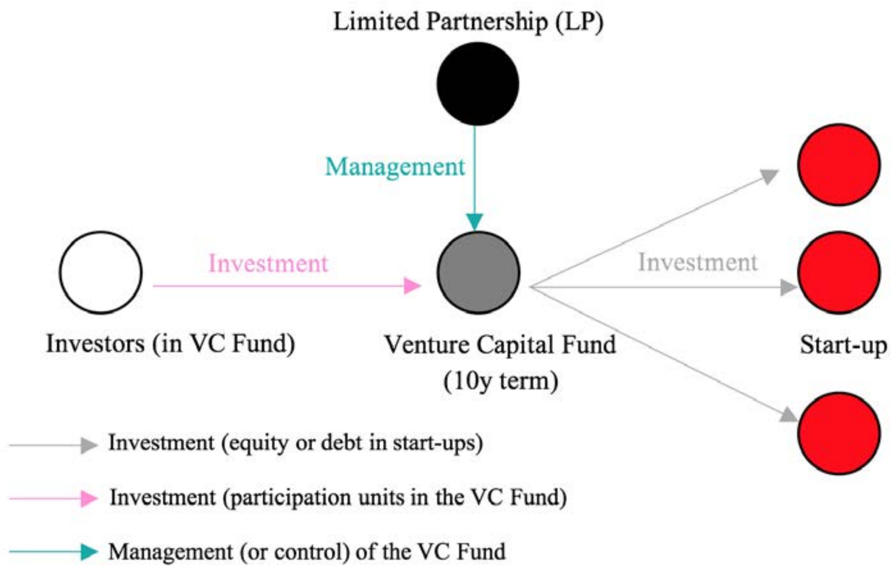
World: An Empirical Analysis of Venture Capital Contract") [15]. These authors analyse 200 investment in venture capital of 118 target companies of 14 investment firms. The data source was the investment firms themselves.

[31]Most literature assumes these 2 scenarios, forgetting the most likely one (bankruptcy or loss of all capital invested) or intermediate ones (return lower than expected adjusted to risk).

[32]There were no IPO's (Initial Public Offer) in the Portuguese Stock Market Exchange (PSI20) between 2006 and the period under analysis (end 2015). Since mid-2014 the PSI-20 has fewer than 20 companies, which are highlighted the exits of Banco Espírito Santo (2014), Espírito Santo Financial Group (2014), Banif, (2015, which was excluded from trading after another bank resolution) as well as Cimpor, Brisa, Impresa and Teixeira Duarte. On the other hand, in 2016 entered Sonae Capital, Corticeira Amorim and Montepio, but they still have less than 20 companies of listed companies.

[33]Also, linking with the (initial) uncertainty (no information prior about the outcome), presenting as a (multi stage) process with incomplete information.

Figure 6: Investment (and control) flows



The study or analysis that is done to a network has as main objective to identify and to understand the behaviour of the system that originated the network [18] [19][20][21][22].

2 Methodological approach

2.1 Data Collection

First, there was no official systematised and structured data on the type of contracts carried out, or the instruments used, since the information is scarce and contradictory (removing some references in case studies). Several data it was extracted (seeking to use reliable sources) to build a report on the ecosystem of this type of investment.

Considering that this type of events can also be observed through other external signs, namely through the mandatory publications (i.e. the validity of certain events depends on their externalisation, e.g. publication on Companies House, the annual report of VC funds). Instead of relying on a case basis, data was retrieved, collected, treated to present a reasonable sample of:

1. Companies financed through FINOVA[34]

[34]FINOVA – Fundo de Apoio ao Financiamento á Inovação – was established by Decree- Law no. 175/2008, of August 26, as a privileged instrument for the achievement of the objectives established in SAFPRI (Sistema de Apoio ao Financiamento e Partilha de Risco). This program, created under the Quadro de Referência Estratégico Nacional (QREN), whose financing entities are the Programa Operacional Factores de Competitivi-

2. Ministry of Justice (MJ) – "Portal das publicações [35]

3. "Crunchbase"[36] and webpages[37]

4. Annual Reports (financial statements)[38]

Data extracted and used per subject (entity):

- **Target Companies**: MJ (Companies House), "Crunchbase", Annual Reports, WebPages and FINOVA (including, COMPETE and PME Investimento).

---

dade (COMPETE) and the Programas Operacionais Regionais de Lisboa e Algarve, aimed to promote the dissemination of financing instruments that provide better Financing conditions for Portuguese SMEs. Source: `http://www.pmeinvestimentos.pt/fundos-sob-gestao/finova/apresentacao/` Data collected and processed in CSV: `https://d-vf.github.io/finova/` (consolidated version "Fusão FINOVA por fases")

[35]Ministério da Justiça - Portal da publicaçõeses. Under art. 167.º of the Commercial Registry Code, (heading "publicações obrigatórias"), defines that they are compulsory publications, on a public access website regulated by an ordinance of the Minister of Justice, thus referring to Decree-Law no. 129/98 , of 13 May, with subsequent amendments, in which art. 6.º (Collective Persons) provides for the following acts and facts relating to legal persons to be registered in the RNPC: a) Incorporation; b) Change of name or designation; c) Change of object or share capital; d) Change of the location of the registered office or postal address, including the transfer of the registered office to and from Portugal; e) The change of the economic activity code (CAE/NACE); f) Merger, spin-off or transformation; g) cessation of activity;h) Dissolution, liquidation or, termination of the return to activity. It was extracted from the records relating to 175 companies referenced as an object of investment (own site of venture capital companies; PME Investimento e Crunchbase), for a total of 1179 events were analysed. After analysis, 754 relevant occurrences were validated (e.g., the same fact described "update" can either have a mere correction of a name, or as an amendment to the articles of incorporation, with only the last one being validated). From these validated records, the relevant and systematically reported elements were extracted into classes, namely: share capital; increase, number of shares (or "quotas"), nominal value, object, Economic activity (CAE) (in which they were extracted through the fiscal number "NIF"), type of fact reported; Date (assuming that of the publication as it appears in the register in the Ministry of Justice, modality and form of subscription, categories if they exist, the number of ordinary shares and the other classes referred thereto, if referenced, were recorded. Bonds. Data (clean and structured) in csv `https://d-vf.github.io/raw-companies-house-records/` Consolidated version direct link: `https://d-vf.github.io/raw-data-fusion-en/`.The last update was in 29/09/2016.

[36]Crunchbase database available at https://www.crunchbase.com/

[37]The following were extracted: 1. Webpages - information on the various websites of funds and management firms about entities subject to investment; 2. Crunchbase - Financing Rounds reported on Crunchbase (of funds based in Portugal). Link to data processed in csv: `https://d-vf.github.io/raw-crunchbase-and-webpages/`

[38]Were extracted, data relating to related parties, subsidiaries, investments in venture capital, taken from the annual reports of various entities that have interests in VC (or in companies' subject to investment). Preferred to the latest available accounts (2015), in which only when it was nonexistent details of the subsidiaries referred to, if appealed to older documents (e.g. Investment Fund). In the case of entities with financial closing at a different date (not 31 December 2015), those of March 2016 were considered (case of SICAR, Luxembourg). Description, Entity (information issuer) and Source are referenced in each document: `https://d-vf.github.io/raw-relatorios-de-contas-act-nov-2016/`

- **Investors (Funds/FCR's)** : Web pages, CMVM, FINOVA (including, COMPETE and PME Investimento) and Annual Reports.

2.2 Data Fusion

After treatment and data structuring individually (by source), it was consolidated all the collected information.

Resolution of conflicts between sources

There are several methods for resolving conflicts from heterogeneous data sources, where deleting records, would compromise the sample. Considering the structure and reliability of the different sources. "MJ" and "PME Investimento" are official sources and are more relevant than the others; the annual reports are audited accounts by (official) third parties. Finally, "Crunchbase" depends on (usually) investors completing the data and, using press releases or other signals that depend, both parties, wanting to promote or not the financing round.

Most cases were solved by considering that only had 1 investor and was verified the dates of investment (or rounds). Thus, it resorted to different sources. The initial sample is composed of 180 companies (175 national and 5 foreign, e.g. although there is a company established in Portugal, the whole investment or, at a certain point is carried out in a company where there is a relationship of domain or group, or have the same shareholders and even trademarks and, it was preferred to add those and not allocate that investment to the domestic company).

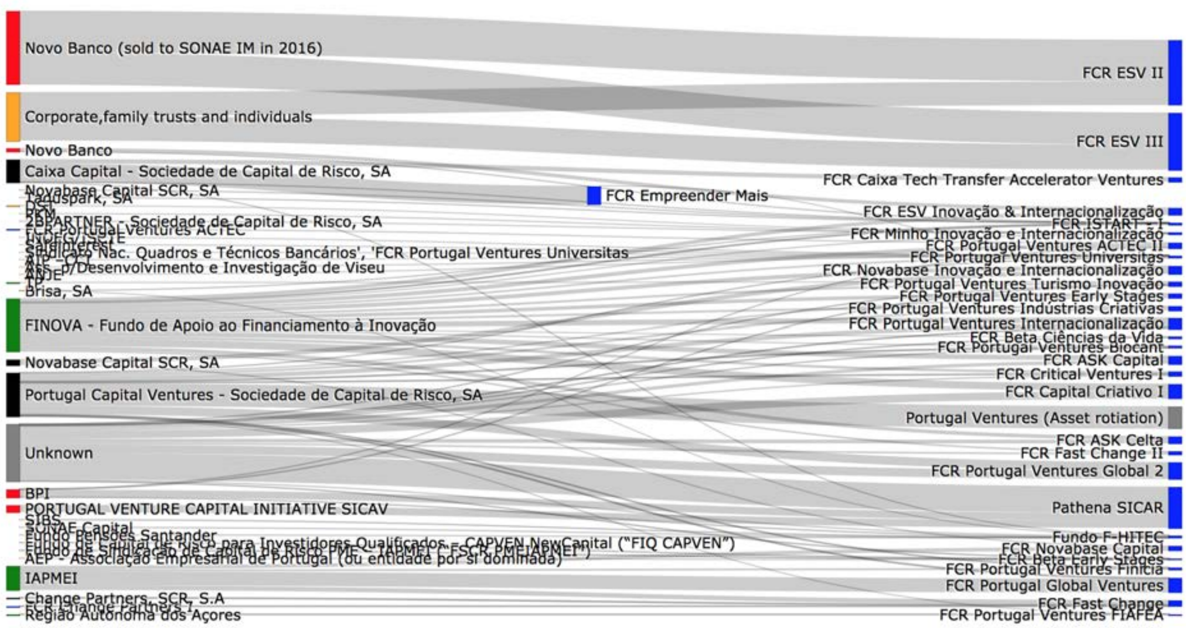
Considering the data of the sample (investment forms used in Portugal), it was added the foreign companies that invested in VC Funds, as well as others who have invested in those target companies. It was also collected the incorporation or constitution dates, as well as the start date of the relationship of the vectors (x) and the end date (y).

2.3 Investors

If was not known the closure of the entity[39], it was assumed that it was $[x, +\infty]$, knowing it was of $[x, y]$. The same applies to the edges.

Represented in Figure 7 infra, there are the investors (left side) in Venture Capital Funds[40] i.e., subscribed investment units of the VC funds (on the right side), as the analysis is on the VC investment done by Venture Capital Funds, whereas the same Managing Entity can hold several funds (with different investment policies and amounts) under management.

Figure 7: Sankey Diagram (VC funds endowments)

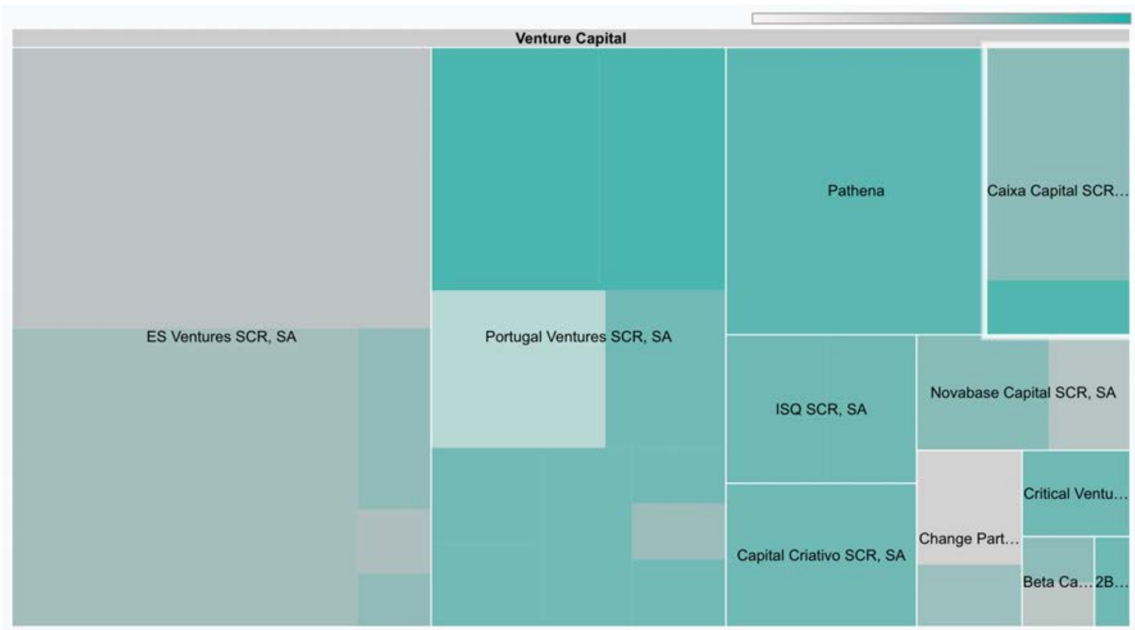


ES Ventures (Funds were sold in 2016) and Pathena are those that managed funds with larger endowments, especially above 30 M€, between the band of 15-20 M€ stand out Portugal Ventures, Caixa Capital and Capital Criativo. The remaining have most values below 10 M€. Thus, considering only the funds mentioned above, there is not only a predominance of funds managed by Portugal Ventures, but also a part of them are co-financed (through public instruments, notably within Compete (FINOVA).

[39]Funds registered at CMVM + Portugal Venture Capital initiative SICAV (PVCi SICAV). Registered with the CMVM (in accordance with Regulation CMVM n. o 3/2015, Regulation on Venture Capital, Social Entrepreneurship and Specialised Alternative Investment) and previous regulation. Were considered 30 venture capital funds that have in their investment policy, invest in the segment "venture capital". The total value of VC Funds was 486.15 M(€value of subscription).

[40]Cf. Table on the Annex. 1. VC Funds, per Management Entity, Amount and incorporation year.

Figure 8: Tree map (Management Entity per amount under management)



Considering the total number (and the value of subscription) of all funds[41] were considered 30 venture capital funds[42] that have in their investment policy, investment in the segment "venture capital". By calculating the Herfindahl-Hirschman Index (HHI), it is possible to affirm, on the one hand, it is a very concentrated industry (close to 2500[43] and, Portugal Ventures and Espírito Santo Ventures (now Armilar Ventures) are key players. It is therefore considered that the total value of VC Funds was 486.15 M€.

[41] The area corresponds to the amounts under management (486.15 M€) of funds by investment management firm and the colour graduation (from grey to green), the years in which they were constituted, where grey corresponds to the oldest and green to newest).

[42] Registered with the CMVM (in accordance with Regulation CMVM n.º 3/2015, Regulation on Venture Capital, Social Entrepreneurship and Specialised Alternative Investment) and previous regulation.

[43] Using the Guidelines of the US Department of Justice [23], available at: `https://www.justice.gov/atr/herfindahl-hirschman-index`)

2.4 Network Structure

It was mapped [44] the agents (Investors and Target Companies) as the contracts (investment flows) between these two over time, by:

Table 1

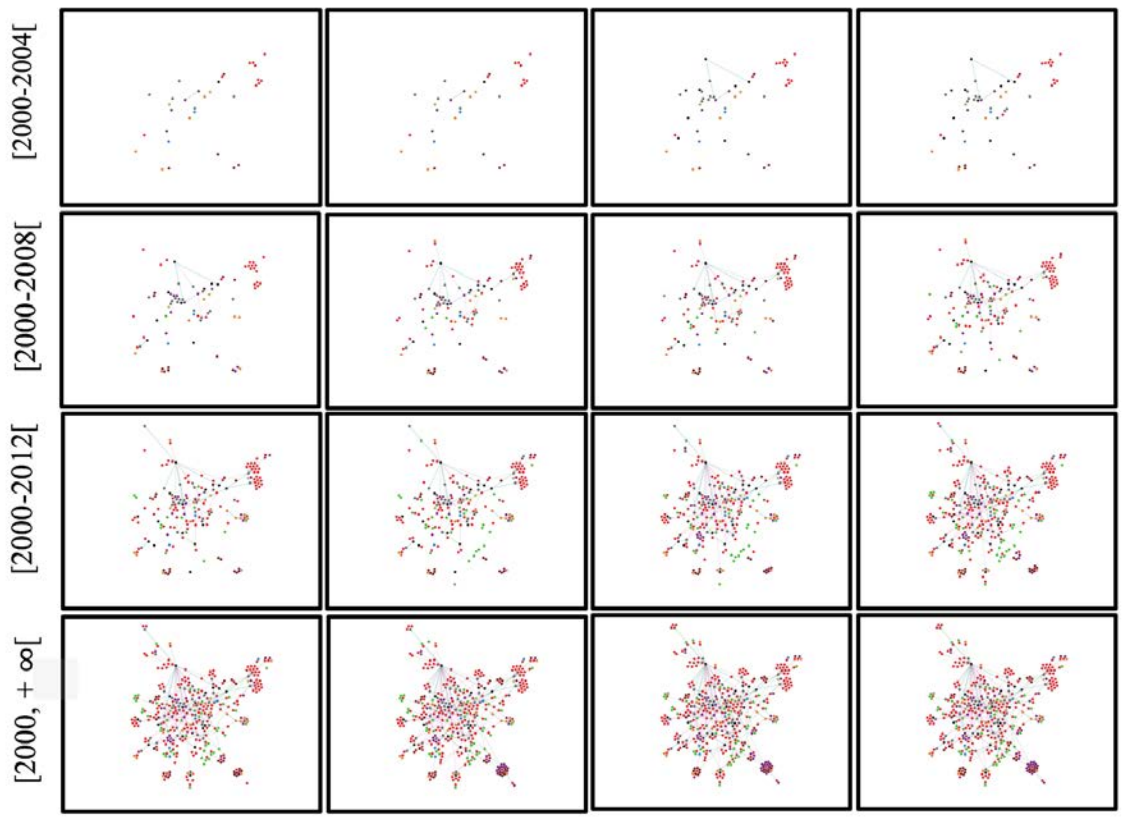
	Absolute number	Percentage	Colour
Nodes	**424**	**100%**	
Target Companies	206	48,58%	FF0000
VC Funds	50	11,79%	666666
Limited Partners (LPs)/SCR's	18	4,24%	000000
Limited Partners (LPs) - Outside PT	30	7,07%	800000
Special Purpose Vehicles (SPV)'s	49	11,55%	33CC00
Corporate	22	5,18%	FF8000
Universities	5	1,17%	0080FF
Banks	3	0,07%	BCBD22
Other	41	9,66%	800080
Edges (directed)	**595**	**100%**	
Investment in start-ups	451	75,79%	999999
VC Fundraising	91	15,29%	E377C2
Management entity (or controlling entity)	53	8,90%	008080

[44] It was done with the software "Gephi" version 0.9.1. An interactive version can also be explored through this link: https://graphcommons.com/graphs/ee67705e-797d-4019-8104-522cf8a61b61?show=analysis-cluster Data files (.gephi and .csv) are available at: https://github.com/d-vf/RAW-Network-Graph

3 Ecosystem Visualisation (2000-2015)

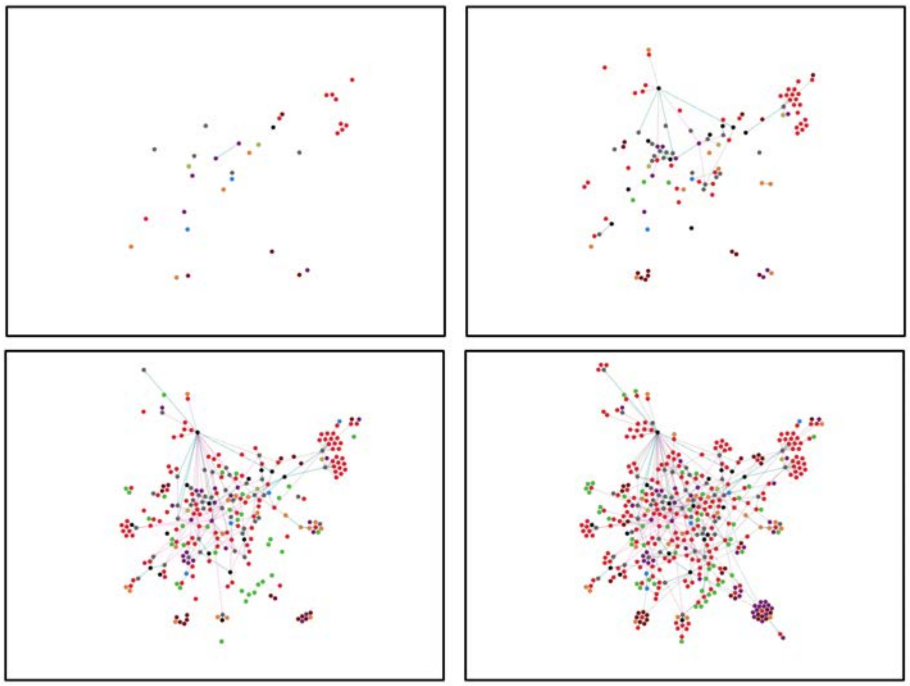
Considering subsequent time intervals of 1 year, the following figures show the evolution of the ecosystem, starting in 2000 until the end of 2015.

Figure 9: Network evolution 2000-2015



Considering the evolution of the number of nodes as edges, for 5-year intervals:

Figure 10: Evolution per 5 years' intervals

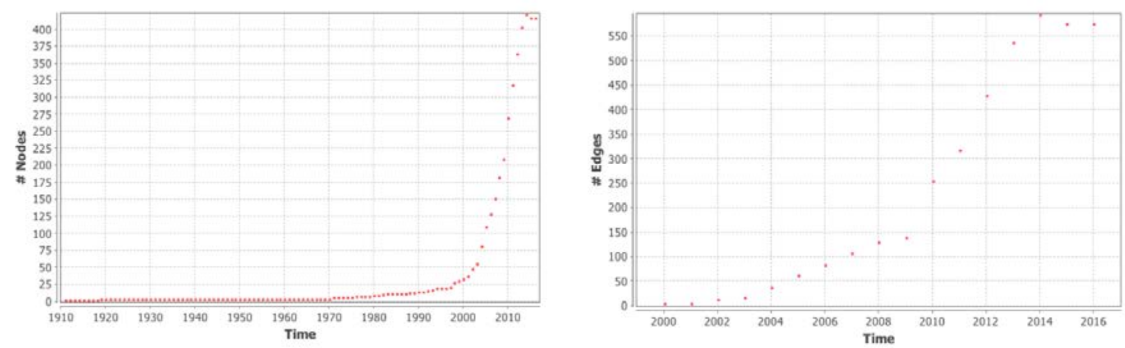


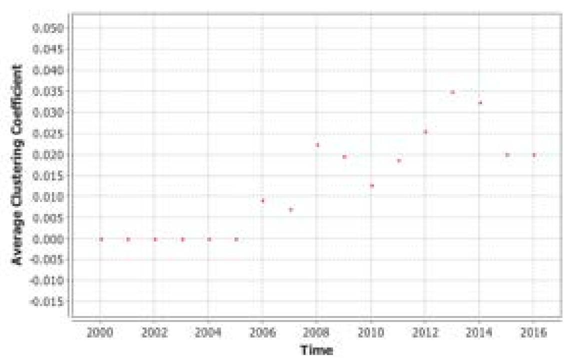
Dynamic Network Metrics

3.1 Statistical measures

The measures can be divided according to the type of analysis that is to be carried out, that is, if the analysis is done at the level of a node, allowing to discover its importance in the network in general or, if the analysis is done at the level of the own network, identifying global network structure characteristics and behaviour that generated it.

Figure 11: Nodes (right) and Edges (left) Times Series





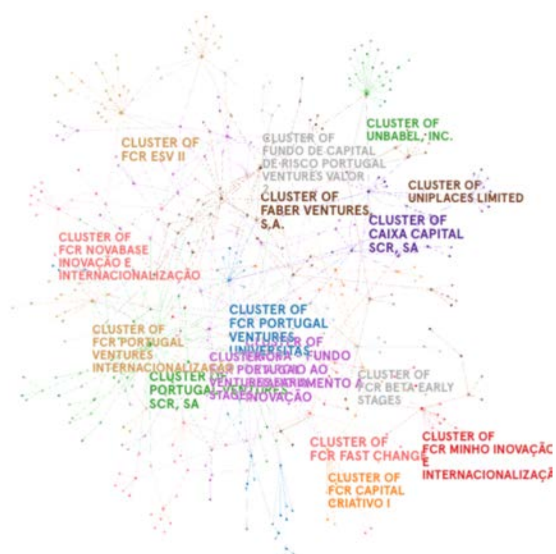


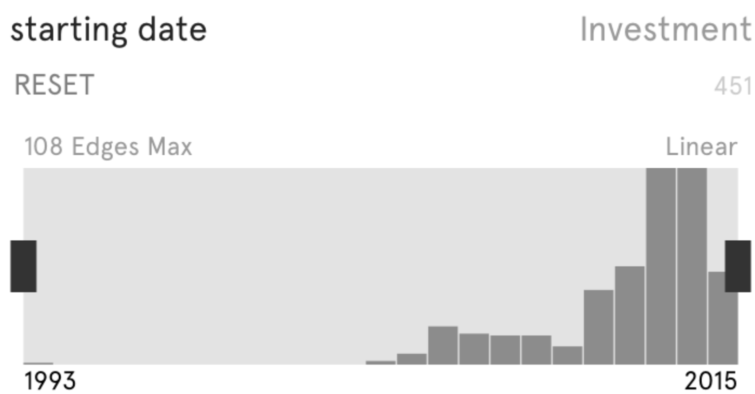
Figure 12: Average Clustering Coefficient

Figure 13: Main Clusters

4 Findings

- It is visible in the graphs, the nodes relative to VC Funds have more connections (edges), than the target companies (which are only recipients of funds). Between 2000 and 2005, no major changes are visible, in which there are some clusters, namely Portugal Ventures (former PME Capital), and Change Partners;
- Between 2005 and 2010 it is possible to see an (initial) increase of interconnections, as well of actors, maintaining the growth trend until 2015;
- The higher the coefficient of agglomeration over time, the greater the increase of the overall ecosystem, still looking to the average node clustering, it is clear, that VC funds, have a higher coefficient than targeted companies (less number of connected nodes);
- If there was a renewal (the case under analysis), this coefficient would be more or less constant. This data is in line with the amounts under management in venture capital reported by the CMVM, where this phenomenon is also visible;

Figure 14: Investment (starting date) 452, max. 108 Edges



- Increases are visible, both in the number actors (nodes) and the number of links (edges) over time (namely VC fundraising in 2011 and VC investment in 2013 and 2014). On the other hand, there is no renewal, i.e., if the edge is maintained it is because there was no divestment (the link remains);
- The nodes with the highest relative importance are, as expected, Portugal Ventures and FINOVA. Considering that they are directed edges, if these two nodes are removed, a more substantial number of the target companies would not have financing. As it is visible, the existing communities have as node connector (where they pass more edges) the VC Funds and their managing entities.

Figure 15: Portugal Ventures (right) and FINOVA (left) Clusters

4.1 Validation and limitations

- Compared with the CMVM aggregate data (which sums participations, per fund), as already analysed in 2015 there would be 720 participations in Venture Capital, whereas an entity reports it, the same company (object) being the subject of co-investment, will have all these reported holdings (of the same company). On the other hand, one of the reasons for cross-referencing and using different sources was, not privileging those that are reported by the parties;

- It was not looked for investments before 2006. The last data verification and update it was in September 2016, when no investments or follow up tranches were considered afterwards;

- The present work is limited to venture capital investment operations, namely through Portuguese venture capital funds and having as target companies, national companies.

5 Conclusion and Future Work

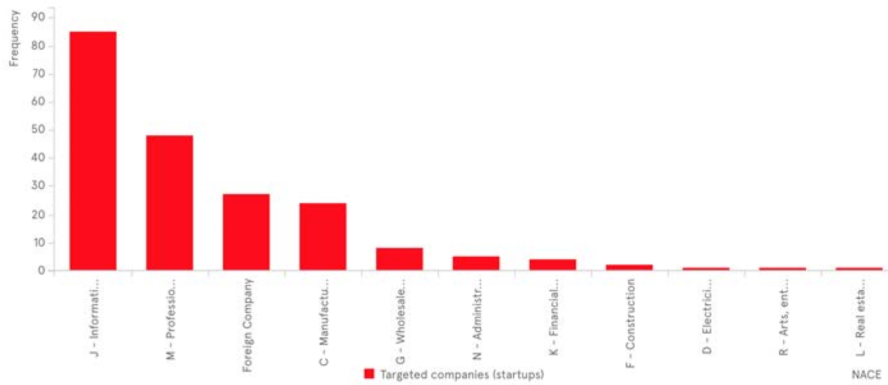
- Temporal networks, could be used namely when looking at Venture Capital, namely to investment flows (more than the amount raised, should be looked until the exit of such investments);

- If the network is dependent on few nodes (resilience and its vulnerability, i.e. dependency of public funding), the relative importance of such cases should be considered;

- The overall ecosystem and its structure seem a better indicator than the annual aggregate accounts. Start-ups, as highly dependent on VC funding and, on another hand, the need to exit after a certain period (usually 10 years) of the VC Fund, should be looked to flows across time (namely if the VC Funds are in an investment or disinvestment periods);

- Other metrics could be used, namely related to the relative position of a given node (as betweenness, closeness, eigenvector), so as the structure and features of the hubs (modularity). Lastly, adding other attributes could improve the understanding of the ecosystem (as NACE[45], origin, among others).

Figure 16: NACE of the target Companies (Start-ups)



[45]Nomenclature Statistique des activités économiques dans la Communauté européenne (NACE) or Statistical Classification of Economic Activities in the European Community

6 Annex

Table 2: VC Funds, per Management Entity, amount and incorporation year (Fig 7. Sankey Diagram Fig. 8 Tree map)

Management Entity	**Fund**	**Amount (M) €**	**Year**
ES Ventures SCR, SA	FCR ESV II	88.00	2006
ES Ventures SCR, SA	FCR ESV III	77.00	2009
ES Ventures SCR, SA	FCR ESV Inovação Internacionalização	10.00	2011
ES Ventures SCR, SA	FCR ISTART - I	3.00	2011
Portugal Ventures SCR, SA	FCR Portugal Ventures Universitas	3.72	2011
Portugal Ventures SCR, SA	FCR Portugal Ventures Early Stages	7.17	2011
Portugal Ventures SCR, SA	FCR Portugal Ventures ACTEC II	7.78	2011
Portugal Ventures SCR, SA	FCR Portugal Ventures Indústrias Criativas	6.00	2011
Portugal Ventures SCR, SA	FCR Portugal Ventures Turismo Inovação	6.00	2011
Portugal Ventures SCR, SA	FCR Portugal Ventures Internacionalização	14.22	2011
Portugal Ventures SCR, SA	FCR Portugal Ventures Biocant	4.00	2011
Critical Ventures SCR, SA	FCR Critical Ventures I	7.00	2011
Change Partners SCR, SA	FCR Fast Change	9.00	2003
Change Partners SCR, SA	FCR Fast Change II	5.00	2011
Beta Capital SCR, SA	FCR Beta Ciências da Vida	2.50	2011
2BPARTNER SCR, SA	FCR Minho Inovação e Internacionalização	4.00	2011
ISQ SCR, SA	FCR ASK Celta	10.00	2011
ISQ SCR, SA	FCR ASK Capital	11.00	2011
Novabase Capital SCR, SA	FCR Novabase Inovação e Internacionalização	11.36	2011
Novabase Capital SCR, SA	FCR Novabase Capital	7.14	2005
Caixa Capital SCR, SA	FCR Empreender Mais	25.00	2009
Caixa Capital SCR, SA	FCR Caixa Tech Transfer Accelerator Ventures	6.00	2015
Capital Criativo SCR, SA	FCR Capital Criativo I	20.50	2011
ES Ventures SCR, SA	FCR F-HITEC	3.50	2008
Beta Capital SCR, SA	FCR Beta Early Stages	2.50	2005
Pathena	Pathena SICAR	55.80	2012
Portugal Ventures SCR, SA	FCR Portugal Global Ventures	20.37	1999
Portugal Ventures SCR, SA	Portugal Ventures Global 2	23.18	2015
Portugal Ventures SCR, SA	Portugal Ventures Finicia	4.00	2007
Portugal Ventures SCR, SA	Portugal Ventures FIAEA	1.10	2011
Portugal Ventures SCR, SA	assets rotation	30.30	2015
Total		**486.15**	

References

[1] NASA, *Technology readiness level definitions*, https://www.nasa.gov/pdf/458490main_TRL_Definitions.pdf, Accessed:23-06-2018.

[2] J. Smith, R. Smith, and R. Bliss, *Entrepreneurial Finance: Strategy, Valuation, and Deal Structure*. Stanford University Press, 2011, ISBN: 9780804777582. [Online]. Available: https://books.google.pt/books?id=xKW4LM3JL-QC.

[3] M. Black; Fischer; Jensen Michael C.; Scholes, *The Capital Asset Pricing Model: Some Empirical Tests*. New York: Praeger, 1972.

[4] F. H. Knight, *Risk, Uncertainty and Profit*. Boston, MA: Houghton Mifflin Co, 1921. [Online]. Available: http://www.econlib.org/library/Knight/knRUP.html.

[5] H. Markowitz, "Portfolio selection", *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, Mar. 1952. DOI: 10.1111/j.1540-6261.1952.tb01525.x.

[6] A. Damodaran, *Price and value: Discerning the difference*, http://people.stern.nyu.edu/adamodar/pdfiles/country/valueversuspriceNew.pdf, Accessed:23-06-2018, 2018.

[7] O. Hart and J. Moore, "Foundations of incomplete contracts", *Review of Economic Studies*, vol. 66, no. 1, pp. 115–138, 1998, available at http://www.nber.org/papers/w6726.

[8] O. Hart, "Financial contracting", *SSRN Electronic Journal*, vol. 39, no. 4, 2001. DOI: 10.2139/ssrn.269241.

[9] P. Aghion and P. Bolton, "An incomplete contracts approach to financial contracting", *The Review of Economic Studies*, vol. 59, no. 3, p. 473, Jul. 1992. DOI: 10.2307/2297860.

[10] D. L. Meadows and C. of Rome, *The limits to growth*. New York: Universe Books, 1972.

[11] A. Damodaran, "Valuing young, start-up and growth companies: Estimation issues and valuation challenges", *SSRN Electronic Journal*, 2009. DOI: 10.2139/ssrn.1418687.

[12] J. Lerner and A. Schoar, "The illiquidity puzzle: Theory and evidence from private equity", *SSRN Electronic Journal*, 2002. DOI: 10.2139/ssrn.322420.

[13] E. Schneider and F. Y. Edgeworth, "Mathematical psychics. an essay on the application of mathematics to the moral sciences.", *Economica*, vol. 2, no. 6, p. 235, Mar. 1935. DOI: 10.2307/2548685.

[14] B. Zider, "How venture capital works", *Harvard Business Review*, no. November-December, pp. 131–139, 1998. [Online]. Available: https://hbr.org/1998/11/how-venture-capital-works.

[15] S. N. Kaplan and P. J. Strömberg, "Financial contracting theory meets the real world: An empirical analysis of venture capital contracts", *SSRN Electronic Journal*, 2000. DOI: 10.2139/ssrn.218175.

[16] Arthur, W. Brian and Steven N. Durlauf and Lane, David A., "The economy as an evolving complex system ii", *Proceedings Volume XXVII,. Santa Fe Institute Studies in the Science of Complexity*, no. November-December, pp. 15–449, 1997.

[17] W. B. Arthur, *Complexity economics: A different framework for economic thought*, https://www.santafe.edu/research/results/working-papers/complexity-economics-a-different-framework-for-eco, Accessed:30 04 2018", 2013.

[18] A. Albert, R.; Barabási, "Statistical mechanics of complex networks", *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002.

[19] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. V. Alstyne, "SOCIAL SCIENCE: Computational social science", *Science*, vol. 323, no. 5915, pp. 721–723, Feb. 2009. DOI: 10.1126/science.1167742.

[20] M. E. J. Newman, "The structure and function of complex networks", *SIAM Review*, vol. 45, no. 2, pp. 167–256, Jan. 2003. DOI: 10.1137/s003614450342480.

[21] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks", *Physical Review E*, vol. 69, no. 2, Feb. 2004. DOI: 10.1103/physreve.69.026113.

[22] H. A. Simon, "The architecture of complexity", *Proceedings of the American Philosophical Society*, vol. 106, no. 6, pp. 467–482, Mar. 1962. DOI: 10.1111/j.1540-6261.1952.tb01525.x.

[23] U. D. of Justice; Federal Trade Commission, *Horizontal merger guidelines (08/19/2010)*, https://www.justice.gov/atr/horizontal-merger-guidelines-08192010#5c, Accessed:01 07 2018, 2010.

How Rankings Go Wrong: Structural Bias in Common Ranking Systems Viewed as Complex Systems
work in progress

Patrick Grim,* Natalia Jenuwine, Jared Stolove, Adrian Apaza,
Hannah vanWingen, Jaikishan Prasad, Paulina Knoblock, Callum Hutchinson,
Chengxi Li, Kyle Fitzpatrick, Chang Xu, Catherine Ming

Center for the Study of Complex Systems, University of Michigan,
1085 S University Ave, Ann Arbor, MI 48109, USA

*To whom correspondence should be addressed; E-mail: pgrim@umich.edu.

We introduce agent-based techniques to analyze inherent structural bias in abstract models of common ranking systems such as PageRank, HITs, and Reddit. In the complex dynamics of reputational loops, an element's ranking itself influences factors in terms of which rank is calculated, resulting in the amplification of divergence and the exaggeration of small random and path-dependent differences. Agent-based models of basic algorithms employed in PageRank, HITs, and Reddit are constructed. The hope is that with further development such models will allow comparisons of bias dynamics and effects across a number of parameters.

Introduction

In many real-world examples, such as college ranking and online search, objects with very similar quality can end up with significantly different rank (*1–4*). Sometimes the data that a ranking system relies on may be dubious (*5*). But even in the best of conditions and with the cleanest data input, we argue, the very structure of some familiar ranking systems can result in distorted informational output (*6*).

The basic idea of all familiar ranking systems is an attempt to read objective quality — what sites, papers, or posts are genuinely worth reading — from social measures of what is read and responded to by whom. In any such system there will be a looping factor: a site, paper, or post will be widely read because it is ranked highly, but will be ranked highly precisely because it is widely read. How severe the loop — and how far rank will come unglued from quality — will depend on the degree to which users base their choice of a site, a paper, or a post on rank as opposed to some independent judgment of quality.

The extent to which looping constitutes a distorting factor, however, will also differ with different ranking algorithms. What we offer here is an initial review of results regarding approaches with an incomplete sample of results regarding three familiar algorithms — PageRank, HITS and Reddit. This remains a work in progress: a more complete paper with a fuller development of techniques and results will appear elsewhere.

PageRank

PageRank analyzes the structure of links between websites in order to determine their ranking (*7–8*). Using a network structure, PageRank treats each website as a node and hyperlinks between websites as directed links between the nodes. Each site is initially assigned a rank value of 1 divided by the total number of nodes. At each time interval, PageRank divides the rank value for each node by the number of outgoing links from that node, and this value is sent as an incoming value to the node at the other end of each outgoing link. Each node's value is then replaced by the sum of its incoming values. PageRank also redistributes an extremely small amount of value equally between the nodes regardless of links, representing the possibility of an individual typing a URL directly into the search bar rather than clicking a link. PageRank then ranks the sites in order of greatest to least node value.

We construct a series of simulations in which pages are assigned an inherent 'quality' between 1 and 100. At each step in the evolution of the model, a small number of pages are added to the network — much as pages are progressively added on the internet, and roughly as nodes are added in a preferential attachment network — and then each page may create additional links to other pages, with probabilities based on the receiving page's inherent quality and current rank.

As noted, the basic idea of all such ranking systems is an attempt to read objective quality from social measures. In the model we can track how well rank corresponds to inherent quality at different settings — in particular, as a function of the degree to which links are formed in terms of a measure of inherent quality or in terms of already-established rank.

The relationship between rank and quality changes significantly as the proportion of link formation determined by rank increases. Fig. 1 tracks this relationship at four different levels of the influence given to rank. When links are determined by quality alone, as shown in the upper left panel, the most highly-ranked pages all have relatively high quality. While many high-quality pages end up with low rank, low-quality pages cannot receive rank above a certain quantity. Also, the spread in rank between pages is moderate, with the most successful pages receiving around 3.5 times the average amount of rank. As the influence of rank on link for-

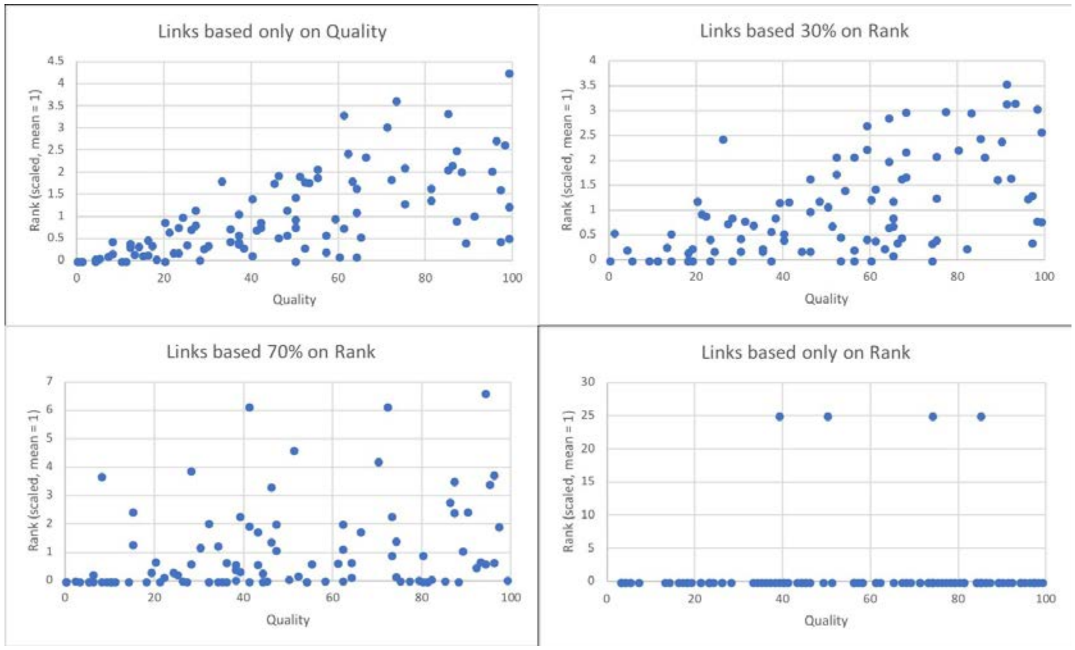


Figure 1: Rank vs. quality in PageRank as the probability of link formation is calculated in terms of quality alone, rank 30%, 70% and entirely in terms of rank. Rescaling of the y-axis should be noted.

mation is increased, as shown in the following panels, lower quality pages are able to obtain a higher rank; fewer pages have high rank; and the difference between the highest and lowest ranked pages becomes much larger.

We introduce 'discrepancy' as a measure for divergence of rank from quality. Discrepancy represents the root mean squared error of rank as a measure for quality, where both are scaled to sum to 1 across all pages. The mean and standard deviation of discrepancy as more weight is given to rank are shown in Fig. 2. Discrepancy increases first gradually and then rapidly as the proportion of influence given to rank approaches 1. The standard deviation also increases, suggesting that when users base decisions strongly on rank, the system's ability to approximate quality is subject to large random variation.

We also track average quality of a page linked to — that is, average quality of the receiving page across all links — as a proxy for the quality that a surfer of the resulting web structure might experience. This measure is around 65 when links are based mainly on quality, and starts decreasing once weight given to rank exceeds around 50% (Fig. 3). With links based only on rank, it reaches 50 — what we would expect if surfers selected pages randomly.

As agents respond more to rank — which we target as the more realistic case — the discrepancy between rank and quality is higher, fewer pages dominate the network, and the average quality of a page linked to is lower.

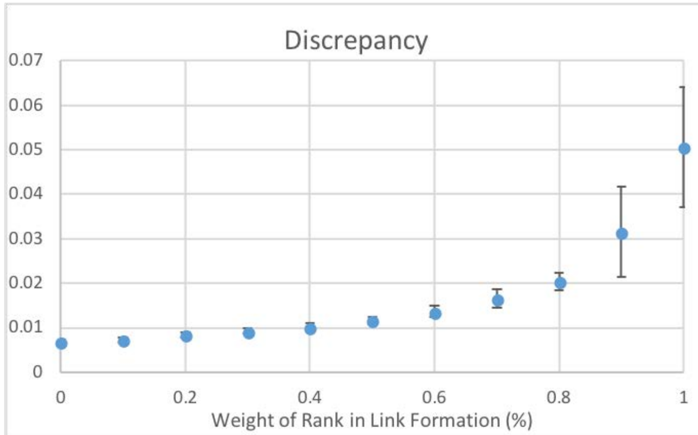


Figure 2: Weight of attention to rank as opposed to quality in link formation and the resultant discrepancy between modelled rank and quality of PageRank sites.

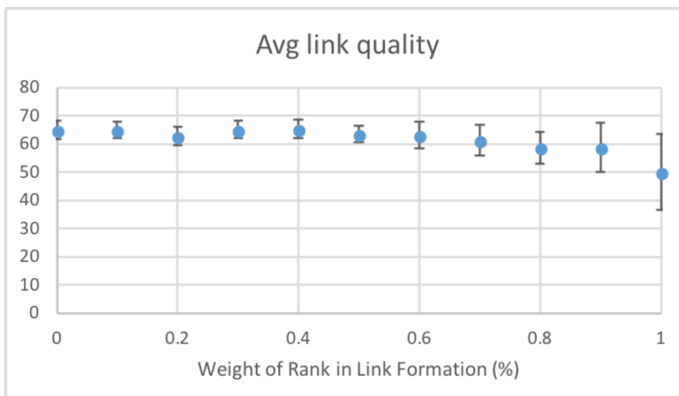


Figure 3: Weight of attention to rank as opposed to quality in link formation and the resultant average quality of the page receiving each link.

HITS

The HITS Algorithm (*9–10*) is used by Academia.edu, a social networking website that shares papers and monitors their impact. In that academic instantiation, the algorithm assigns values to papers and to authors in mutual recursion. Papers are assigned PaperRanks on the basis of the number of recommendations they receive from authors, weighted by the AuthorRank of those authors. AuthorRank is calculated on the basis of the PaperRanks of that author's own papers. This instantiation differs from a more general version of HITS, in which the equivalent of AuthorRank is calculated using the PaperRank of the papers an author recommends rather than those they wrote.

In our model, both papers are assigned a constant built-in 'quality value.' During setup, authors decide which papers to recommend based solely on paper quality. Subsequently, papers are scored on the prestige and number of authors that have recommended them, while authors are given AuthorRanks on the basis of the PaperRanks of their papers. Each iteration, authors recommend new papers with probabilities based on each paper's inherent quality and current PaperRank.

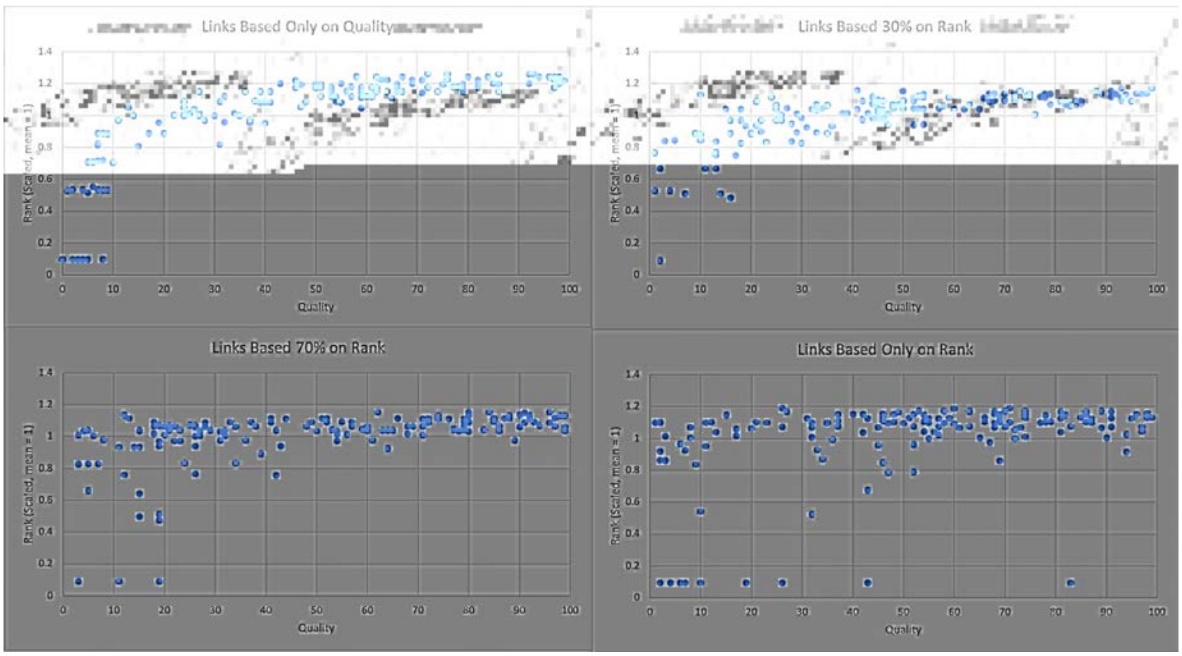


Figure 4: Rank vs. quality in HITS as the probability of link formation is calculated in terms of quality alone, rank 30%, 70% and entirely in terms of rank.

Here, as in the case of the PageRank model, we can track how well PaperRank corresponds to a paper's inherent quality at different levels of influence given to rank (Fig. 4). Again it is clear, as expected, that the relationship between rank and quality decays as modeled users attend increasingly to rank. In other regards comparison of Fig. 1 and Fig. 4 shows that the algorithms behave quite differently. PageRank allocates a majority of rank to a handful of highly successful pages. The degree of inequality increases as users act based increasingly on rank, indicated by the change in y-axis scaling in Fig. 1. HITS distributes rank more evenly between many papers, and the overall distribution does not change when users attend to rank, allowing for the same scaling throughout. However, as the level of rank-influence increases, rank in HITS tends to approach the mean for most papers, regardless of quality, while more low-quality papers rise to the top and more high-quality papers drop below the average.

Reddit

Reddit.com is a social news aggregation site where users can discuss and rate posted content. Content is distributed by topic among various 'subreddits' where posts with the highest net rankings (upvotes minus downvotes) rise to the top of the page. By ranking based on net score, Reddit is intended to create a preference toward non-controversial content, since a post which receives 50 upvotes and zero downvotes will be ranked the same as a post with 500 upvotes and 450 downvotes. Comments on posts are also voted upon and change order correspondingly. Reddit's main Front Page shows posts with the most upvotes across all subreddits, with the order of links changing constantly based both on the time of submission and user votes. A post's score will not decrease as time passes but newer posts will get a higher score.

Our model of Reddit simulates website users reading posts and voting on them. We model users as deciding to read a post and to give it an upvote or downvote based on two factors: its objective 'quality' (a number between 0 and 100) and its rank as determined by the Reddit algorithm. Each modeled user is assigned a threshold that determines how likely they are to read, upvote, or downvote a given post on average. During the simulation, users periodically 'leave' the site, replaced by users with different thresholds. New posts with assigned 'qualities' are periodically created as well, then ranked according to Reddit's algorithm.

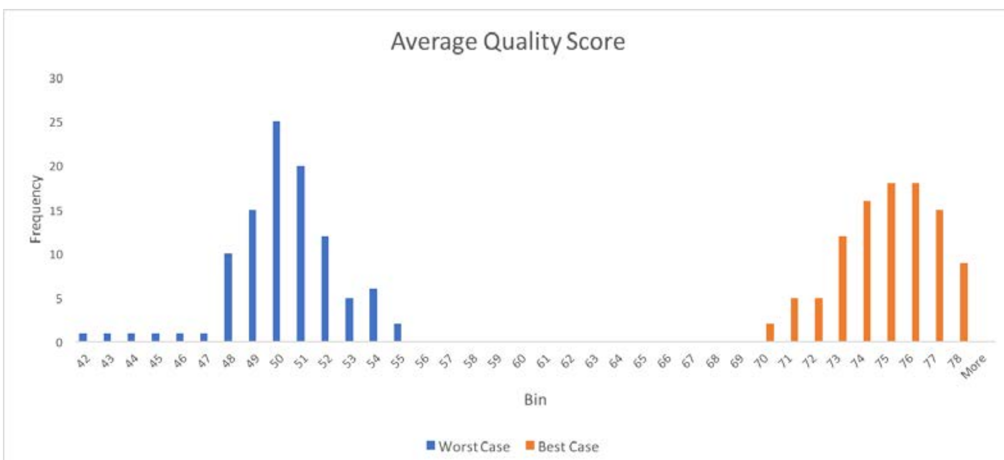


Figure 5: The distribution of average quality of a read post, or 'average quality score,' in both the best- and worst-case scenarios over 100 runs of the model.

In order to get a baseline measure to help us evaluate the algorithm, we start off with two highly unrealistic situations: a 'best case scenario' for both reading and voting and a 'worst case scenario.' The best-case scenario considers a situation in which users are able to tell the objective quality of a post before even reading it, and make their decisions to read and vote based solely on quality. The worst-case scenario describes a situation in which users read and vote on posts based solely in terms of rank. In that case rank is totally unrelated to quality. The goal of any ranking algorithm is to offer users posts of high quality. We evaluate how well Reddit performs by an ‘average quality score’ — the average quality of a post read by a user during the simulation. Fig. 5 shows the distribution of the average quality score over 100 instances of the simulation, under both the worst case and best-case scenarios. The worst-case scenario depicts an average quality score distributed approximately normally around 50, suggesting that user’s reading habits are totally unrelated to quality. In contrast, the best-case scenario depicts a distribution centered at a quality score of 75, the approximate mean value of user’s reading thresholds.

We then evaluate how Reddit performs given users who base their reading and voting more on quality or more on rank. Fig. 6 shows an evaluation of the Reddit algorithm’s robustness to rank-bias based on how close the average score is to the best case scenario when we introduce varying levels at which reading and voting are based on rank.

		Voting Bias					
		0	0.2	0.4	0.6	0.8	1
Reading Bias	0	74.5	74.399	74.774	74.26	72.844	72.33
	0.2	73.51	72.84	73.477	74.62	73.94	71.78
	0.4	73.73	72.21	75.16	71.22	72.17	70.82
	0.6	70.97	71.057	72.105	71.033	71.466	68.788
	0.8	69.561	68.34	68.865	67.068	70.044	60.78
	1	68.3827	66.83	65.77	65.98	63.01	49.025

Figure 6: The average quality of a post read in versions of the model with varying levels of Reading Bias and Voting Bias. Cells that are relatively closer to the best case scenario are colored green, those relatively closer to the worst case scenario are colored red.

In our model, performs quite well even at relatively high levels of both kinds of bias. Even with both bias parameters set to .6 rank, the resulting average quality score is 71.033, much closer to the ideal scenario than to the worst-case scenario.

Reddit’s robustness in the face of rank-related bias appears to be due to the built-in penalization of older posts in the ranking algorithm. Reddit’s algorithm provides newer posts with an advantage in two ways. First, newer posts are given a one-time increase in their initial raw score, called the ‘time damp.’ Second, additional upvotes only impact a post’s score logarithmically, so highly ranked posts must acquire exponentially more votes to counteract the advantage given to newer posts. PageRank, we’ve noted in passing, is dominated by a small number of pages. Reddit’s built-in penalization of older posts avoids this consequence. This is a reflection of the differing goals of Reddit and PageRank: the former attempts to show users fresh content while the latter focuses only on quality. Reddit’s methodology can reduce the impact of path-dependent behavior by preventing dominant posts from emerging unfairly, but it does so at the cost of pushing high-quality pages down as well.

The Looping Effect and Bias in Familiar Ranking Systems

The basic idea of all familiar ranking systems is an attempt to read objective quality — what sites, papers, or posts are genuinely worth reading — from social measures of what is read and responded to by whom. In any such system there will be a looping factor: a site, paper, or post is widely read because it is ranked highly, but it is ranked highly precisely because it is widely read. The existence of loops in ranking systems appears to be ubiquitous and inevitable: its role is demonstrable in even simple agent-based models of PageRank, HITS, and Reddit.

As noted, what we offer here is a review of work in progress. It is clear in all three models that rankings become poorer indicators of quality as users increasingly base their judgments on rank. In PageRank and Reddit we have calibrated this in terms of the decreasing quality of content seen by users. But it is also clear that all three models can withstand moderate amounts of rank-based bias, showing most significant divergence of rank from quality when rank is weighted at 80% or more. Further comparison between the three systems given the current measures is difficult, given differences in both structure and informational goals — Reddit's emphasis on temporally fresh content, for example. We can, however, advance several hypotheses regarding early results that will guide us in further work. Some of the differences between Reddit and the other algorithms may be due to its explicit emphasis on newer elements, dampening an inherent reputational looping bias toward older posts evident in the other two algorithms. Our initial results indicate important differences in the patterns of relationship between rank and quality at different settings. Here a partial explanation may lie in the fact that HITS uses two values in mutual recursion, effectively doubling the effect of reputational loops. Confirmation and expansion of initial results as well as further exploration of our suggested hypotheses remain as tasks for further work.

References and Notes

1. M. Gladwell, The Order of Things: What College Rankings Really Tell Us. *New Yorker Magazine, Feb. 14 & 21* (2011).

2. M. Bastedo, N. Bowman, The U.S. News and World Report College Rankings: Modeling Institutional Effects on Organizational Reputation. *American Journal of Education* **116** , 163-184 (2010).

3. M. Bastedo, N. Bowman, College Rankings as an Interorganizational Dependency: Establishing the Foundation for Strategic and Institutional Accounts. *Research in Higher Education* **52** , 3-23 (2011).

4. M. Glenski and T. Weninger, Rating Effects on Social News Posts and Comments. *ACM Transactions on Intelligent Systems and Technology* **8**, 10.1145/2963104 (2016).

5. C. O'Neal, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown, New York, 2016).

6. A. Nematzadeh, G. Ciampaglia, F. Meneczer, A. Flammini, How algorithmic popularity bias hinders or promotes quality. arXiv:1707.00574v2 [cs.CY] (2017).

7. S. Brin, L. Page, The Anatomy of a Large-Scale Hypertaxtual Web Search Engine. *Computer Networks and ISDN Systems* **30 (1-7)** , 107-117 (1998).

8. L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web (Stanford InfoLab, 1999).

9. J. Kleinberg, Authoritative sources in a hyperlinked environment. *Journal of the ACM* **46 (5)** , 604-632 (1999).

10. J. Kleinberg, Hubs, Authorities and Communities. *ACM Computing Surveys (CSUR)* **31**, article 5 (1999).

Introduction to Decision Process Theory

Gerald H. Thomas[1*]

[1]Milwaukee School of Engineering, Milwaukee, WI 53202, USA
*E-mail: thomas@msoe.edu.

We believe that decisions are not random but are a connected network rippling from the past into the future. Is it not commonly accepted that human decision making is the result of free will and is therefore not deterministic? Yet there are opposing views, such as systems dynamics, game theory and complex network theory that some aspects are deterministic. When this view is extended to include the system dynamics of space and time, an engineering approach to decision-based social structures becomes possible, with potentially profound consequences. This talk describes our approach to pulling these ideas together into a unified theory based on the mathematics of differential geometry.

Introduction

Recently I taught a course for junior and senior engineering students with the idea to present a theory of differential game theory and provide the students the ability to apply such a theory to decision problems in areas such as economics. The presentation for this conference is based in part on that course and in part on a recent book on decision process theory, as well as other published work including some preliminary concepts that we presented at the NECSI 2004 Conference (*1*). There, we explained the theoretical ideas that may be new to some, though they rely on concepts that are well-known each in their own areas from physics, including string theory and gravity, to game theory and differential geometry. We believe it is the application of these dynamical ideas to the field of game theory that is new.

We note the difficulties students have using ideas based on differential geometry because they are not familiar with the computational techniques that are commonly associated with them. In the course, Wolfram Mathematica was used to help focus attention on the ideas so as not to be unduly distracted by the computations. This was successful with the students as they turned their focus on problems they were familiar with and applied the ideas to those problems. This approach may be more generally successful, which is a part of the current project.

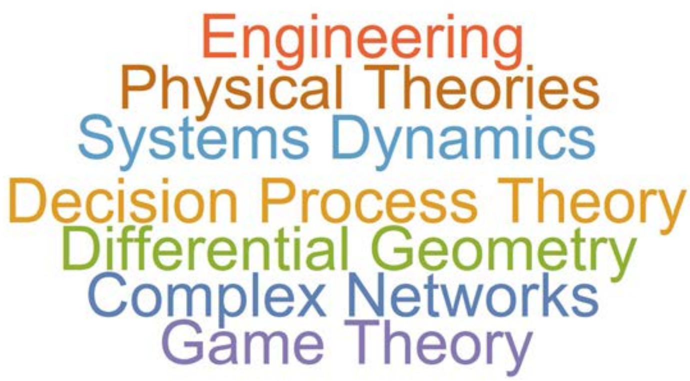


Figure 1: Multiple strands.

There are multiple strands (Fig. 1) that led us to decision process theory: *e.g.* the theoretical side consists of physical theories, *(2-4)*, differential geometry *(5-6)* and game theory *(7-10)*. This theoretical side also encompass an empirical side, which we supplement with ideas from complex networks *(11)*, engineering and systems dynamics, *e.g. (12-14)*, with additional cues from business and strategic planning, *e.g. (15)*.

There are three key theoretical ideas for decision process theory: from game theory, the idea that the mixed strategies are causal behaviors even though the actual decisions humans make are not deterministic; from complex networks, that mixed strategies may be connected in a spatial network; and from differential geometry the idea of creating a unified space-time using the space of mixed strategies and the game theory notion of utility as a measure. It takes some effort to justify these statements *(16)*. For this talk we take these ideas as given and look to their consequences.

The equations from differential geometry can be written as a geodesic equation, where we separate out the aspects we wish to deal with directly from those that need to be included but where an average treatment is acceptable:

$$R_{\mu\nu} - {}^{1\!}/_{2} g_{\mu\nu} R = -\kappa T_{\mu\nu}.$$

They provide a complete systems dynamics description of the behaviors in the theory as both functions of space and time in terms of the metric $g_{\mu\nu}$ and its related curvature components $R_{\mu\nu}$.

They differ from the equations from physics at a foundational level: the indices span time and points in the space of mixed strategies as used in game theory. They share the look of general relativity theories by separating the geometry constraint aspects from the cultural fluid aspects defined by the conserved energy momentum tensor $T_{\mu\nu}$. They build on the ideas of Kaluza *(3)* and Klein *(4)* by clearly identifying the symmetries of the metric and reformulating the geometry accordingly, an approach borrowed from string theorists, such as *(17)*. These partial differential equations are challenging to solve numerically, though we believe possible.

One important consequence of this formulation is that local symmetries (such as an isometry along direction k) break the general metric into component parts, one of which, ${F^k}_{ab}$, can be identified with the game theory notion of payoff for a player and justify the identification of the metric with utility. This isometry identifies the strategy k with the game theory notion of player. The concept generalizes to the idea of a code of conduct adopted by multiple players. Thus, this differential game theory diverges conceptually from game theory.

From an empirical view, we take the perspective of systems dynamics as opposed to the Bayesian view taken by many in game theory. We look to identify local mechanisms that can be directly related to observation and use the theory to provide the systems dynamics closed loops in both time and space to generate the global behavior of these interconnecting mechanisms. This approach has a lot in common with engineering studies. Not only does the mathematics identify ${F^k}_{ab}$ to the payoff for a player, the same mathematics identifies this structure as an analog of the electromagnetic field. This allows many of the ideas from engineer-ing and systems dynamics to be carried over. It is closely related to one common-ality observed in business situations: the notion of steady-state behaviors and how such behaviors are modified by an external stimulus *(12)*.

The theory establishes relationships between areas that have been well studied in business (operational dynamics), engineering and game theory and pulls them together into a common framework. It suggests new and insightful ways for expanding the vocabulary of the dynamics of decision making. It was with this goal that a course was presented to juniors and seniors from electrical, computer and software engineering programs to see how they might create and apply such ideas.

An outgrowth of this exercise was an investigation into the meaning of player ownership, which extends the published works *(18)*. We describe some highlights of this exercise, which serves to illustrate the theory, how it might be applied and our current research. The engineering students were assumed to have no knowledge of game theory, differential geometry and would have limited understanding of physical theories outside of what they might need in their area. For example, software engineers would not know much about electromagnetic fields. However, it was assumed that all would have been exposed to a standard engineering curriculum that includes calculus, differential equations, and basic science including how to apply the scientific principle. All would know some basic statistics. It was assumed that they were not familiar with Mathematica or the Wolfram Programming Language, which formed a common starting point to introduce them to the required basics for the theory.

Strategic Flows

Strategic flows provide a direct connection to game theory as flows are proportional to the rate at which a player chooses an *intensive* strategy. Game theory distinguishes intensive and extensive strategies. A nice introduction to students is the game of tic-tac-toe. We are all clear that the play of the game differs from the well-known fact that one can always win or at worst tie the game. Looking at the game in this way identifies the intensive strategy or choice that can be made, which is different from the individual moves one makes while playing.

Game theory dictates that the static equilibrium can be deduced from knowing the relative utility of the different intensive choices and is summarized by the game matrix for the two players. When students were asked to pick a game, by far the most popular was the game of chicken. Here is a game matrix for a form of chicken in which each player believes they are rewarded if they crash and are not rewarded as much if they swerve:

$$G_k = \begin{pmatrix} 80 & 60 \\ 0 & 20 \end{pmatrix}$$

Students start with this form of the game and are guided through the analysis, which means identifying the Nash equilibrium(s); in this game, there are two. Students are then introduced to the identification of this game with the corresponding payoffs ${F^k}_{ab}$ in decision process theory. The game theory connection, see *e.g. (19)*, is also made to symmetric games and linear programming, which provides a useful tool here as well.

The verbal analysis of the game, which the students are asked to provide, is that if both ”crash”, one player, say ”blue” gets a huge reward. This means they are fearless; they enjoy the thrill and don’t mind losing. If both ”swerve”, there is some reward based on depriving the other player the joy of crashing. If ”blue” chooses to ”crash” and ”red” swerves, this is better, but not as much fun as both crashing. If ”blue” swerves and ”red” crashes, ”blue” gets nothing from this: there is no crash and no reward at having deprived ”red” the joy of crashing. ”Blue” chooses to crash and assumes that ”red” chooses to ”swerve”. Of course, ”red” sees the same story so also chooses to ”crash”.

Game theory determines that this game has two Nash equilibriums. Each player sees the view that the best choice is for them to ”crash” and assumes that the other player will ”swerve”. What is interesting is that a decision process theory model yields a composite strategy that is a mixed strategy favoring ”swerve” to crash by a 11:2 ratio. We don’t take the ratio too seriously, but believe the game illustrates a depressing option, since the only way to counter a ”crazy” is to be a ”crazy”: we view the choices made here to be crazy. We call this model the ”flip chicken” game, since it ”flips” the behavior of each player fearing the ”crash”, to each player enjoying the ”crash”.

Decision process theory views this game as more than a statement about equilibrium but a statement about the behavior around the equilibrium. The structure of the behavior around equilibrium provides insight into the forces that maintain that equilibrium. The equations for the flow follow directly from the conservation laws based on the field equations from the theory,

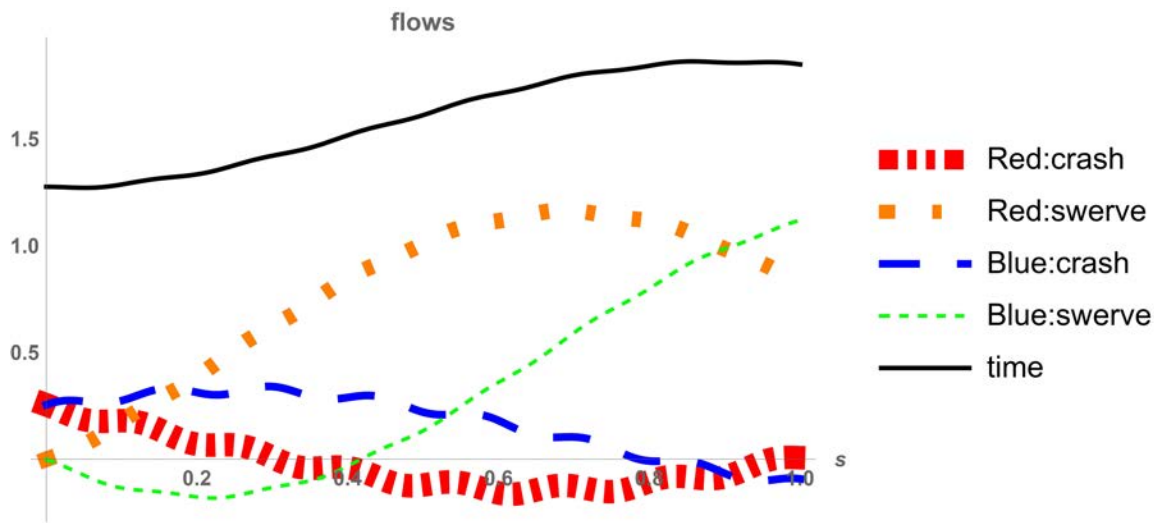


Figure 2: The flows for a game of chicken with no ownership constraints. This game of ”flip chicken” is different from the usual one in that each player sees a benefit to crashing based on the game matrix.

which requires a specification of the cultural context $T_{\mu\nu}$.

When the cultural context can be ignored, the equations that result are geodesics in the extended space, which if we focus only on the active strategies (those not associated with a specific isometry), are:

$$g_{ab}\frac{DV^b}{\partial\tau} = V_k F^k{}_{ab} V^b$$

These are the flow equations, along the streamline of proper time τ, as modified by the antisymmetric payoff field. These are purely the consequences of the differential geometry of the game theory, though to some of you experts, they may remind you of the relativistic version of the flow of a charged fluid in an electromagnetic field. The expected result would be harmonic oscillations around the magnetic field direction, which in this case would be along the composite Nash equilibrium of the players.

We generate the structure by disrupting the equilibrium, defined as the null vector of the above flow equation at some initial instant. We then track the behavior that results. For this model and a typical disruption, we get `(Fig. 2)`. The flows oscillate around the composite equilibrium position and demonstrate that in this theory, there are harmonic forces that become visible once a disturbance has been applied.

For a one-quarter course, this is about as far as we were able to get, given the need to bring everyone up to speed in the necessary game theory, differential geometry theory, and the requisite Mathematica language needed to understand and appreciate the result. However, getting to this point was a big accomplishment for the students. When challenged for their final project to come up with their own examples and explain the resulting harmonic curves, they provided surprising understanding of the concepts with innovate examples.

Network: Consequences of Ownership

A fuller appreciation of the theory becomes clear with full solutions to the differential geometry field equations that include the cultural contexts. It is a challenge to obtain such solutions. However, based on the insights gleaned from one set of complete numerical solutions *(16)*, we have been recently studying solutions that include ownership, where any given strategy is owned by only one of the players. Our view is that ownership is a result of the cultural context. This provides a non-trivial example of the tensor $T_{\mu\nu}$:

$$T_{\mu\nu} = \mu V_\mu V_\nu - \sum_{J \in players} \varphi_J Q[J]_\mu{}^\lambda Q[J]_\nu{}^\rho$$

The sum extends over each of the players (defined by the corresponding isometry) with a projection operator $Q[J]_\mu{}^\nu$ that in the appropriate frame of reference, is diagonal and unity for that players strategies and zero otherwise. This provides a precise definition of ownership. The energy flows $V_\mu = \{V_k, V_a\}$ for the inactive (isometry) and active strategies, and the energy density are determined by the conservation of longitudinal and transverse momentum given input assumptions on the contextual couplings φ_J.

Ownership is treated as a constraint, analogous to constraints in mechanics (such as a perfect fluid) where the energy and momentum are included but not treated as dynamic variables. We are familiar with treating fluids in terms of their pressure and energy density as opposed to a detailed molecular description in terms of a vast number of variables. The consequence is that instead of considering the flow along a streamline as we did above, for two players we consider two distinct streamlines, one for each player. Along the diagonal, we may recover the single streamline case, though this depends on the initial conditions established on the boundaries.

Though we won't go into depth here nor did we go in depth in the course, we are currently pursuing the network aspects of the theory by looking into the consequences of this ownership model. This allows us to make visible our understanding of strategic flow as the energy momentum flow of the system (not treated as an analogy but defined precisely in the differential geometry). We recover many of the distinctions that were identified in our fuller treatment *(16)*. In addition, we obtain an advantage that this approach can be used by students to study any two-person game with a straightforward generalization to three or more players.

The streamlines we consider are the characteristic curves along the vector flow for each player. For two players, the two characteristic curves are labeled $\left\{ \begin{array}{cc} s_1 & s_2 \end{array} \right\}$. They relate to the total effort r_k each player expends. Such total efforts play no strategic role in classical game theory. Based on our experience from exact solutions, we assume in addition that on the boundary, the inactive flow is $e_k = e_{k0} \tanh {}^u\!/_c$, where $u = r_1 - r_2$: this gradient capacitance

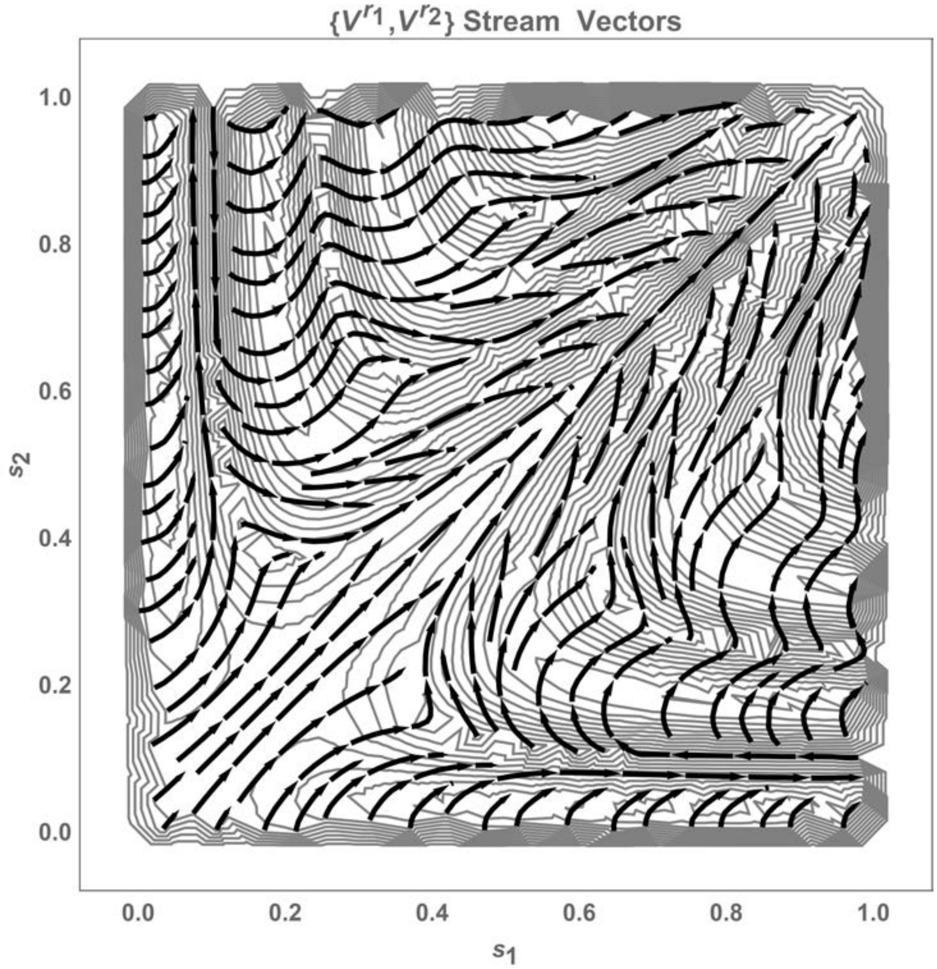


Figure 3: Player stream Vector Flows for a model with ownership and gradient capacity model constraints.

model assumption may give different results for zero-sum and non-zero-sum games and thus, may give useful insights into the structure of decision processes.

In this theory, by moving away from a single direction of flow (along the proper time), we expose the aspects of the theory as seen in `(Fig. 3)`, where we introduce compression and shear with respect to the variation of the flow with (strategic) position. Because of the specific compression assumptions used here, we see that the flows tend towards the diagonal.

Such conditions are imposed on the boundaries and lead to the changes of the behavior that are displayed; the behavior along the diagonal, `(Fig. 4)`, is a dynamic generalization of Nash equilibrium that is quite different from the boundary equilibriums assumptions made at $s_1 = 0$ or $s_2 = 0$. On the boundary $s_k = 0$, for each player k we have assumed each player obeys its Nash equilibrium. Each player chooses to "crash". The result along the diagonal is that each player initially makes this choice, but at larger proper times, the composite strategy

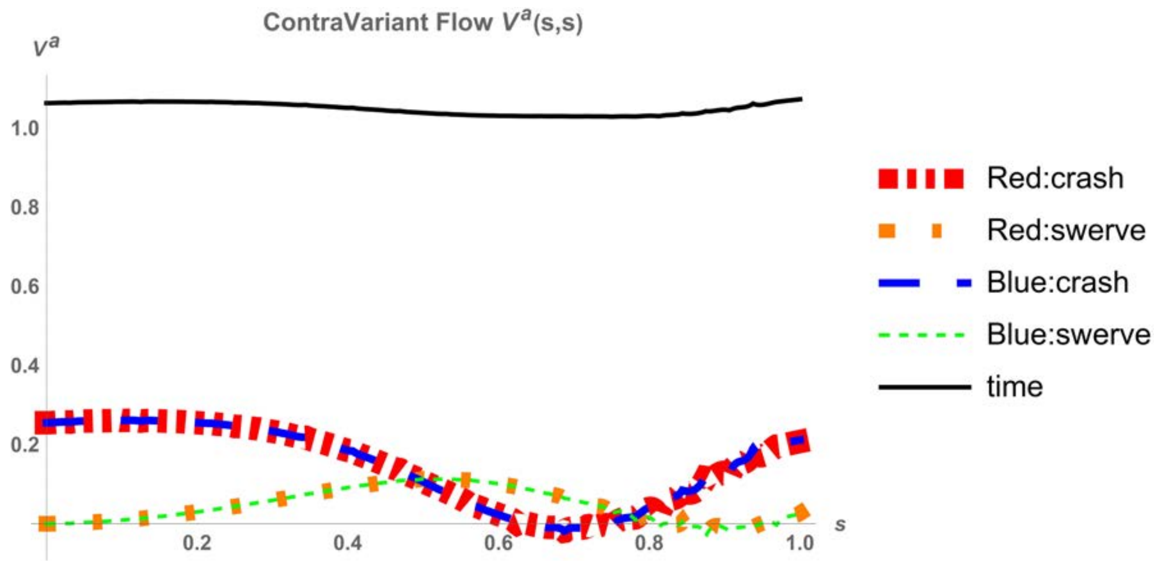


Figure 4: Neutral Behavior of the strategic flows for a model with ownership constraints.

starts to dominate, and each player starts to favor the ”swerve” strategy.

There is equality between the ”crash” flows and ”swerve” flows as well. This is not true away from the diagonal, because of the assumptions made for the shear (which we associate with a player bias towards one or more strategies). Depending on which way one goes, one or the other of the player’s Nash strategy starts to be favored again. Of course, the exact details depend on the initial conditions imposed; what we have identified are general mechanisms that determine the dynamic behaviors.

Conclusions

The theory as first proposed *(1)*, has been published and substantially clarified *(20, 16)*. An infrastructure using Mathematica is now under development *(18)* to quickly compute the consequence of any game and is being used to better understand the concept of ownership. It is also being used to understand other related concepts such as overhead, social context, contextual coupling, competition and bias to name a few. The major challenge is to deal with the underlying non-linear partial differential equations in a way that is easily accessible. Our future goal is to use the same techniques for a more general class of problems solving the full set of field equations of the theory.

Our approach is to produce a series of engineering notebooks that fulfill the initial goal with the ownership models. The notebooks will generate the consequences of any initial set of assumptions about the decision model, which then can be shared. The approach is a generalization to one used with students where we took an engineering approach of summarizing the working formulas needed, with the references provided for those that wanted to dig deeper, along with a

sufficient high-level view of the meaning behind the equations used.
yielding

References and Notes

1. G. H. Thomas and K. Kane, "A dynamic theory of strategic decision making applied to the prisoner's dilemma", in *Unifying themes in complex systems: Vol. VI, Proceedings of the Sixth International Conference on Complex Systems*, A. A. Minai, D. Braha, and B.-Y. Yaneer Eds. (Springer Verlag, 2010), pp. 275-282.
2. R. M. Wald, *General Relativity* (The University of Chicago Press, Chicago, 1984).
3. T. Kaluza, *Sitz. Preuss. Akad. Wiss.* **K1** 966 (1921).
4. O. Klein, *Helv. Phys. Acta. Suppl., IV*, **58** (1956).
5. N. Steenrod, *The topology of fibre bundles* (Princeton University Press, Princeton, 1951).
6. S. W. Hawking and G. F.Ellis, *The large scale structure of space-time* (Cambridge University Press, Cambridge, 1973).
7. J. Von Neumann and O. Morgenstern, *Theory of games and economic behavior* (John Wiley and Sons, Inc., New York, 1944).
8. J. Nash, *The Annals of Mathematics* **54(2)** 286-295 (1951).
9. M. Shubik, *Game Theory in the Social Sciences: concepts and solutions* (MIT Press, Cambridge, 1991).
10. R. B. Myerson, *Game Theory: Analysis of conflict* (Harvard University Press, Cambridge, MA, 1991).
11. A.-L. Barabasi, *Linked* (Plume, a member of the Penguin Group, New York, 2003).
12. D. H. Meadows, D. L. Meadows, J. Randers and W. W. Behrens, *The Limits to Growth (The Club of Rome)* (Signet Book, New York, 1972).
13. P. M. Senge, *The fifth discipline: the art and practice of the learning organization* (Currency Doubleday, New York, 1990).
14. B. Richmond, *An introduction to systems thinking* (iThink@ Software from High Performance Systems, Inc., Hanover, NH, 2001).
15. M. Walton, *The Deming Management Method* (Putnum Publishing Group, New York, 1986).
16. G. H. Thomas, *Geometry, Language and Strategy–The Dynamics of decision processes, Vol. 2 (Vol. 2)* (World Scientific, Singapore, 2016).

17. M. B. Green, J. H. Schwarz and E. Witten, *Superstring Theory (vols. 1 and 2)* (Cambridge University Press, Cambridge, 1987).

18. G. H. Thomas, *Introduction to decision process engineering* (Wolfram Media Inc, Champaign, to be published).

19. R. D. Luce, and H. Raiffa, *Games and Decisions* (Dover Publications, Inc., New York, 1957).

20. G. H. Thomas, *Geometry, Language and Strategy* (World Scientific, New Jersey, 2006).

A compositional lens on the drivers of complexity

Harshal Ganpatrao Hayatnagarkar
Engineering for Research, ThoughtWorks Technologies, Yerwada, Pune, India 411006
harshalh@thoughtworks.com

"What makes a system complex?" **and** ***"Can we compare complexity of one system with another?"*** **are a couple of deep questions in the** ***Complexity science*****. Their answers will enable us to better understand, explain, develop, and manage the complexity of not only human-built systems, but also natural systems. In this paper as an epistemological inquiry and an attempt to answer above questions, I propose that the complexity of a system emerges from interplay of four primary concepts namely scale, diversity, network, and dynamics. Although such ideas have been discussed before, the paper brings in a compositional lens to the discussion along with case studies. It examines the impact of various combinations of these fundamental concepts and the different flavors of complexity emerging out of this amalgam. It then maps these concepts to the existing terminology of complexity science, to understand if these terms can be explained using our compositional lens. For example, complexity as non-linear dynamics in a system emerges from special arrangement of that system's parts, and network properties can help us to further study such arrangements. I hope this lens helps to understand the nature of complexity in systems, and new ways to manage it, which I witnessed in cases of human behavior modeling, software development for big data, and control systems for modern scientific apparatus.**

Introduction

Complexity science provides a scientific philosophical framework to understand systems and phenomena that are called as *Complex*. The term complex has no single agreed upon definition, but it certainly implies difficulty to understand a complete system or phenomenon by understanding its individual parts, and by this semantics much of the World is complex. Even more difficult is to compare complexity of unrelated systems, such as human body, large computer software, economies, and stars. Our particular interest is in characterizing sources of complexity in the sense of challenge involved in effectively achieving various purposes, such as understanding, explaining, engineering and managing. I believe that having an articulated model of the sources of complexity as in understanding *"when a system **becomes** complex"* will enable to better manage the complexity of human-built systems, and perhaps even grapple better with the challenges involved in understanding complex natural systems and phenomena.

Grady Booch famously wrote *"Role of an engineer is to create an illusion of simplicity."* (*Booch(2006)*), and hence from an engineering viewpoint a decent characterization of complexity will help in —1. Understanding a problem in the World, 2. Explaining and communicating this understanding with others, 3. Developing a solution collaboratively, 4. Maintaining/operating it over years.

Complexity science literature has many appealing definitions of complexity, such as those associated *edge of chaos*, *emergence*, *Wolfram's complexity class*, or *non-linear dynamics* as discussed in (*Page(2010)*). A classification of systems as *simple*, *complicated*, and *complex* which is based on *causality*, respectively means independent causal relationships, chained causal relationships, and reciprocal causal relationships via feedback loops (*Edmonds(1999), Page(2010)*). At the same time, certain systems are perceived more complex than others, and this perception could be based on many dimensions. If one can define a measure of complexity across systems, then it provides a guidance for their comparison. Thesis of Dr. Edmonds (*Edmonds(1999)*) is perhaps the most comprehensive survey of, what he terms as, *syntactic measures of complexity*. This work clearly and articulately enumerates these measures, and to an extent discusses their interrelationships. Later, works like (*Kruchten(2012)*) describe *scale*, *diversity*, and *connectivity* as three drivers of complexity. Similarly, my earlier work in the context of monitoring & control system software identified three other drivers namely *scale*, *diversity*, and *evolution* (*Chaudhuri et al(2013)*). In this work, I extend this characterization of complexity along four drivers or dimensions *scale*, *diversity*, *network*, and *dynamics*. I believe that *evolution* is a composition of all four drivers, and is discussed below. Here composition carries semantics as in *interplay*, and not strictly *mathematical composition*; for example, if a system with a million parts of 10 types is more complex than half a million parts of 1000 types. This idea is analogous to how primary colors combine to result into entire color spectrum. In this publication, I believe the novelty comes in concretely presenting this compositional lens using three case studies.

Drivers of complexity

Each of the four drivers has different impact on resultant complexity. They do not exist in isolation, but each may make a system complex differently depending upon its dominance. In this section, I discuss each driver and its semantics.

- **Scale** is the first driver, and mostly implies size, count, or in general magnitude of something. In canonical sense, adding more entities to a system adds to its complexity, such as having 10 entities versus 10 million. So, the scale puts pressure on real-world systems which always have finite resources, and increasing scale pushes system toward deficit.

It can be managed by approaches namely *partitioning*, *summarization*, and *representation* —Partitioning is divide-and-conquer approach; summarization is stating something across large number of entities, such as finding out mean and median of an attribute across a large population; and representation (or sampling) is about substituting entire population (or collection) by one or more entities from it based on some criteria, for example a survey may sample just 1000 people instead of entire population of a million to extend their inference across actual population.

- **Diversity** contributes to complexity by means of more kinds of things. A little increase in diversity may step up complexity by orders of magnitude. For example, *Kolmogorov complexity* of an alphabet with single letter (which means no diversity) is easy to handle, but just addition of one more letter can create strings with infinite complexity via randomness (*Ming and Vitányi(1997)*). *Variety* is also diversity, albeit in a finer sense. For example, we see diverse species in a forest, but also variation within one species (*Page(2010)*). Diversity can be managed by using classification of some kind, for example Linnean classification for species, and in my opinion for variation one needs to use statistical techniques to draw inference.

- **Network** is the third driver and is about arrangement of entities or parts of a system. Arrangement may have some repetitions within, forming a *pattern*. Patterns may indicate presence of deep semantics in them, such as friendship networks, and potential points of failures in an electric grid (*Newman(2003), Costa et al(2011)*). Visualizing a network using many layouts can reveal different insights. Holistically, a network depicts structural relationships in a system.

- **Dynamics** is the last and the only one representing temporality or time aspect. The dynamics comes in to play via network, but also in composition of scale and diversity. Certain configurations such as *feedback loops*, *connected components* contribute in various defining aspects of complexity such as *non-linear dynamics, tipping points, attractors, and path dependence*.

I will now briefly discuss how I perceive the existing concepts in complexity science may be mapped to compositions of the drivers.

- **Emergence** can be seen as systemic network dynamics.

- **Pattern** may be spatial (network) or spatiotemporal (network-dynamical) configuration with repetitions. Some of these patterns become abstractions *atom*, *molecule*, *cell*, *tissue*, and so on, giving a multi-level/multi-scale view (*Vaughn et al(2013)*). In software engineering, patterns which act as abstract standard solutions to recurring problems play an important role in dealing with software complexity (*Gamma(1995)*).

- **Kolmogorov complexity** is a complexity measure accodring to (*Ming and Vitányi(1997)*) and (*Edmonds(1999)*), and I perceive it is composition of *scale*, *diversity*, and *network*. For example, if a string has repeating sequence (*aabbaabb*...) ad infinitum, its Kolmogorov complexity is low, and hence it effectively handles scale. Adding more letters to its alphabet would only decrease positional probability of a letter in a string. However, if one changes a few letters in a long sequence (as network), it may increase the complexity by order of magnitude. It is not computable measure (*Edmonds(1999)*).

- **Self-organized criticality** is a composition of scale, network, and dynamics because a system will hold its criticality up to certain limit (scale) on its self-organized structure (network-dynamics).

- **Tipping point** is a composition of all drivers because a system may be tipped by any of them.

- **Attractor** is a composition of all drivers just like tipping point as it is a regime change.

- **Randomness** can be explained by Kolmogorov Complexity, and in turn by associated drivers. A random sequence of letters has infinite Kolmogorov complexity.

- **Evolution** is a composition of all drivers, since it operates of populations (scale), with intra-species variation and inter-species diversity, via mutations and adaptations (network), over time under selection pressure (dynamics). As discussed in (*Gavina et al(2018)*), we can also see these drivers affecting selection in wild ecology via weak and strong crowding effects.

- **Modularity** implies how a system can be decomposed in to reusable subsystems, that can interface and interact with each other (*Edmonds(1999)*). A large system will benefit if modularized, and each subsystem can then handle its local complexity effectively due to smaller scope, fewer parts, lesser diversity, smaller network, and perhaps resulting into a well-understood dynamics.

*Computational complexit*y, a subdiscipline of *Theoretical Computer Science*, studies resource requirements of an algorithm, such as processing steps and memory. It follows an asymptotic approach, for example complexity $O(n^2)$ is treated equivalent of $O(2 \cdot n^2 + 5 \cdot n + 1000)$. It is a combination of *scale* (depends on number of items to be processed), *diversity* (average, best, and worst case complexity), *network* (algorithm/flowchart is a network), and finally *dynamics* (as exhibited in running the algorithm) characterized by the complexity classes such as subpolynomial, polynomial and non-polynomial.

How to handle the complexity emerging from compositions drivers of the complexity in generic sense is still an open problem. At the same time, three case studies I discuss in subsequent section can reveal some interesting patterns.

Case study: ITER

First case study is of a distributed monitoring & control system (MCS, for short) for a large scientific apparatus namely *International Thermonuclear Experimental Reactor* (ITER), being constructed in France. ITER belongs to a class of systems namely *Cyber-physical Systems* for which hardware and software aspects must be co-developed and pose their own design challenges (*Lee(2008)*). ITER has instrumentation hardware and software elements such as sensors, actuators, controllers, platforms, communication middleware, and so on.

Complexity of ITER puts enormous pressure in ensuring quality of MCS over the long timeline of development and maintenance spanning across decades. Complexity of distributed control systems has been studied before, for example (*Yang(2006)*) discusses complexity of centralization of control, and (*Scattolini(2009)*) discusses complexity of architectures for predictive control. For ITER MCS, initial analysis to understand drivers of its complexity discovered three drivers namely *Scale*, *Diversity*, and *Evolution* (*Chaudhuri et al(2013)*). Here, I take this discussion further by substituting *evolution* with *network* and *dynamics*, and their impact on construction and maintenance of the system.

1. **Scale** manifests itself in the form of few hundred thousand hardware-based and software-based instrumentation elements. Because of their large number, it becomes difficult to keep track of each hardware instrumentation element and its configuration across many software platforms. Entire configuration can easily reach millions of lines of specification source code, which is challenge in its own right.

2. **Diversity** affects the MCS complexity differently for hardware and software. (a) Sensors, actuators, and controllers have diversity as they come in different types, and have variety in their deployment configuration depending upon context. (b) The MCS itself is an integration of various software platforms and components. Each added platform or component, forces the MCS team to address its mapping to every instrumentation element in the system.

3. **Evolution** of hardware and software over decades during and after development poses its own complexity challenge. Hence, the team must ensure to not introduce any defects in the MCS by evolutionary changes. Evolution can be seen as composition of network and dynamics drivers because evolution implies changes in the network over time under selection pressure.

4. **Network** plays its role via hierarchical breakdown structure ITER system in to subsystems and instrumentation elements, which must be connected as per design specification and guidelines. These connections are modeled as a network of components, and then mathematical framework of *graph theory* or *network theory* can be used in analysis and design of control systems (*Chaudhuri et al(2013), Chernyi(2016)*), and in fault analysis (*He and Zhang(2011)*). These elements and their interconnections create a large

number of network paths, making verification of configurations is difficult due to high computational complexity (*Fujiwara and Toida(1982), Abdollahi et al(2016)*), in performing static analysis of the design as a network model (*Kavulya et al(2012), Chaudhuri et al(2013)*).

5. **Dynamics** in the ITER MCS acts through various network paths with feedback loops and decision points. Such networks cannot be analyzed for desired output, and thus may lead to unintended consequences, for example contagion or cascade failures. ITER safety and interlock subsystem keeps monitoring for symptoms of such failures. One more challenge is ITER plasma dynamics, whose shape and temperature cannot be precisely controlled due to many intersecting and high velocity plasma currents.

To address this complexity, ITER MCS is being developed as model-driven control system (*Chaudhuri et al(2013)*). A mathematical model captures details of MCS hardware and software elements, and engineers can specify configuration details using custom-built user interfaces. This model can be checked for known issues and best practices. Finally, code for each platform and component with cross-references is generated using this model. Changes during hardware and software platforms evolution are absorbed by the code generation mechanism, without touching any actual model. A concept called as *Tower of Abstractions (or Tower of Interpreters)* is a language-oriented approach, in which newer abstractions are discovered by users, mainly via reuse (*Wand and Friedman(1988)*), and can significantly aid in complexity management.

Case Study: Big Data

When quoted the term *Big Data*, it is often associated size or magnitude of data. Publication (*Ward and Barker(2013)*) discuses many definitions of the term in a quest to understand its origins and characteristics, and at least two definitions mention *complexity* as an aspect of *big data* —Microsoft's definition is without further clarification on term *complexity* (*Ward and Barker(2013)*), whereas the SAS definition implies data fusion from multiple sources as the measure of complexity (*Gandomi and Haider(2015)*). On the contrary, I find a famous characterization of big data around 5 '*V*'s (initially first 3 '*V*'s) namely —*Volume*, *Variety*, *Velocity*, *Veracity*, and *Value* (*Laney(2001)*) —closer to the compositional lens of complexity without any explicit mention of it. In rest of the section, I discuss the drivers of complexity mapped to *5V* characterization of big data.

1. **Volume** is the size of data, and maps to *scale* driver. Although there is no common benchmark, one can assume 10^{12} bytes (or terabytes) as threshold (*O'Malley(2008)*). Many algorithms with polynomial or more computational complexity struggle to process large volumes of data (*Lelkes(2017)*) (for example, to process n items using an algorithm with cubic time and quadratic space complexities, requires n^3 number of processing steps, and

n^2 storages bytes). Increasing data volume puts significant constraints on system design, development, and usage, and sometimes leaving no solution in sight for typically available resources. As discussed earlier, the strategy to handle data volume involves partitioning, aggregation, and representation depending on context.

2. **Variety** implies structuredness of data, from unstructured to semi-structured to structured. Examples of unstructured data are documents in natural English such as emails and literature; examples of semi-structured data are semantically tagged entities, and object-based/hierarchical documents; and examples of structured data are in the form of relational database tables and programming language source code. Apart from text, images, audio, and video are also data. Thus, variety becomes equivalent of *diversity* driver. In addition, one can perceive that variation of values such as anomalous (outliers), incorrect, and missing values also contribute to complexity by means of uncertainty. To handle diversity in data types, classification techniques such as ontologies and software engineering patterns are useful. In addition, statistical techniques can help to deal with variation in values.

3. **Velocity** in big data is interpreted for speed of data, either inward or outward to a system. For example, if a data is acquired in an online/real-time manner, it has higher velocity than batches at lower frequency. Acquisition, storage, analysis, and visualization are all affected by velocity of data, especially at large volumes (*Chen and Zhang(2014)*), and different approaches are required to handle lower and higher velocity data (*Bonomi et al(2012)*), even though unified processing engines are built (*Carbone et al(2015), Zaharia et al(2016)*). In our context, this velocity is composition of *scale* and *dynamics*.

4. **Veracity** here means accuracy, not in terms of precision but trustworthiness and reputation, because data may have biases, ambiguities, and inaccuracies (*Lukoianova and Rubin(2014)*). Veracity can be established if there is correct and enough metadata linked with data, for example owner, method, origin, date & time, and so on. One can also think of maintaining lineage between raw data, processed data, and insights along with transformations performed at each stage. This creates a network of various linkages of metadata and data together, and presence or absence of such metadata linkages play crucial role in drawing inference from the analyses, and thus can be interpreted as *network* driver of complexity.

5. **Value** is perhaps the most crucial aspect of the big data, and it stands for *business value*. If we look closer, business can see value only if actionable insights can be derived from data analysis. An actionable insight prescribes changes in structure and function of a business to achieve some goal, and thus, evolution of the business over time. Thus, I interpret it as changing the structure (*network*) to impact business *dynamics*. Business dynamics itself can be interpreted for its own complexity drivers but is out of scope for the current context.

Case Study: Human Behavior Modeling

The last case study in this discussion is at the intersection of simulation modeling and behavioral science. Modeling of human behavior is considered as a grand challenge due to its complexity (*Taylor et al(2013), Vaughn et al(2013), Cipresso(2015)*). I witnessed this complexity while composing human behavior models from behavioral fragments (for example, a relationship between two or more behavioral variables) that are grounded in literature of behavioral sciences (*Duggirala et al(2017)*), and in this section I attempt to relate this complexity to the four drivers being discussed.

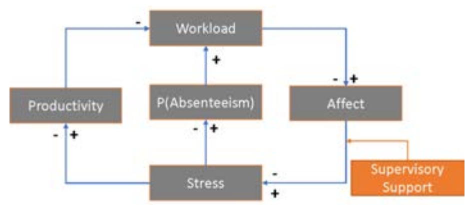
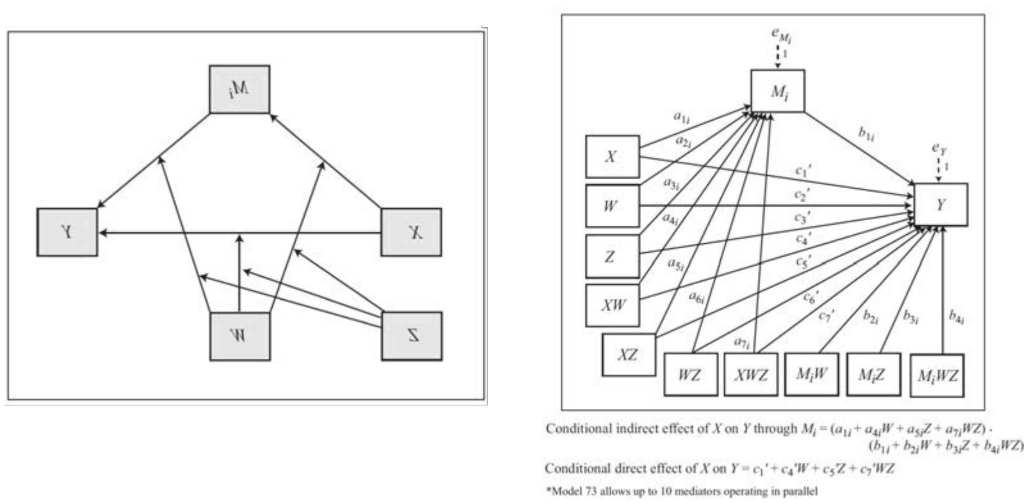


Figure 1: Stress-dynamics model with moderating supervisory support

An example of a composed behavioral model having six variables is shown in 1 that captures stress dynamics of an employee working in a support-services organization, especially to explore role of supervisory support in dealing with stress (*Kumar et al(2017)*). More complex models satisfying mathematical, behavioral, statistical, and computational constraints are possible, but their interpretation becomes difficult (*Duggirala et al(2017), Hayes(2013)*) as depicted in following figures.



(a) Conceptual diagram (b) Statistical diagram

Figure 2: Statistical model template (*Hayes(2013)*) for behavioral modeling

1. **Scale** : Number of variables in a model map to scale. Even a few variables contribute to the complexity due to their standalone and interaction effects as shown in 2 (*Duggirala et al(2017), Hayes(2013)*).

2. **Diversity**: In behavioral sciences, variables are classified in to *traits* such as age and gender, and *states* such as hunger and emotions. In addition, there could be external variables, such as signals, environmental conditions, and so on.

3. **Network**: No behavioral variable works in isolation, and they affect each other via interaction effects. Randomized and controlled studies may reveal a few interactions at a time, and are published in the literature. As discussed in (*Duggirala et al(2017)*), an approach is being formulated to compose models as shown in 1, which is actually a network of variables, based on statistical and psychological primitives of composition such as *mediation*, *moderation*, and their composition as discussed in (*Hayes(2013)*). It is but obvious even one incorrect relationship in such a network-based behavioral model can sway the results substantially. In a different point, a human being is also a network of individual but self-organized cells.

4. **Dynamics**: As discussed for diversity and network, bottom-up behavioral models would clear way for *multi-scale/multi-level modeling* based on emergent abstractions or levels (here I prefer term *multi-level* over *multi-scale* to avoid confusion with *scale* driver). For example, molecular and cellular dynamics in individual biological cells give rise to abstractions such as tissues, organs, individuals, and then all the way up to the society (*Vaughn et al(2013)*). The emergent multi-level view of network and dynamics together helps us to classify variables into *physiological*, *cognitive*, *psychological*, *social*, and *environmental* levels. There is an additional challenge caused by an individual's memory, which makes a lot of behavior *path dependent* based on that individual's life.

As discussed in this section, human behavioral modeling is a grand challenge due to its complexity, and the four drivers can help to characterize this complexity.

Conclusion

As seen from three very different case studies —Monitoring & control software for ITER, big data concepts, and human behavior modeling —the complexity of each can be resolved from a compositional lens, as contributed by the four drivers namely *scale*, *diversity*, *network*, and *dynamics*. Novelty of this work is to describe the lens concretely using these case studies. In the future, I would like to explore epistemology and role of composition in a more nuanced way, such as impact of interactions between two or more drivers on complexity. I also believe it will be useful to contrast complexity of minimalist systems (as in Conway's Game of Life) to real systems (as in natural and human-built systems).

Acknowledgement

I acknowledge guidance of my mentors and colleagues, and thank them in chronological order - Dr. Harrick Vin, Dr. N. Swaminathan, Dr. Madhav Marathe, and Dr. Mayuri Duggirala. I also thank Mr. Santosh Mahale and my other colleagues at Engineering for Research Practice of ThoughtWorks Technologies India for their encouragement and support in getting this work published.

References

Abdollahi et al(2016). Abdollahi A, Pattipati KR, Kodali A, Singh S, Zhang S, Luh PB (2016) Probabilistic graphical models for fault diagnosis in complex systems. In: Principles of Performance and Reliability Modeling and Evaluation, Springer, pp 109–139

Bonomi et al(2012). Bonomi F, Milito R, Zhu J, Addepalli S (2012) Fog computing and its role in the internet of things. In: Proceedings of the first edition of the MCC workshop on Mobile cloud computing, ACM, pp 13–16

Booch(2006). Booch G (2006) Object oriented analysis & design with application. Pearson Education India

Carbone et al(2015). Carbone P, Katsifodimos A, Ewen S, Markl V, Haridi S, Tzoumas K (2015) Apache flink: Stream and batch processing in a single engine. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 36(4)

Chaudhuri et al(2013). Chaudhuri SR, Swami N, Hayatnagarkar HG (2013) Integrated monitoring and control specification environment. In: Proceedings of ICALEPCS 2013, ICALEPCS

Chen and Zhang(2014). Chen CP, Zhang CY (2014) Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences 275:314–347

Chernyi(2016). Chernyi S (2016) Techniques for selecting topology and implementing the distributed control system network. In: IOP Conference Series: Materials Science and Engineering, IOP Publishing, vol 124, p 012048

Cipresso(2015). Cipresso P (2015) Modeling behavior dynamics using computational psychometrics within virtual worlds. Frontiers in Psychology 6:1725, DOI 10.3389/fpsyg.2015.01725, URL https://www.frontiersin.org/article/10.3389/fpsyg.2015.01725

Costa et al(2011). Costa LdF, Oliveira Jr ON, Travieso G, Rodrigues FA, Villas Boas PR, Antiqueira L, Viana MP, Correa Rocha LE (2011) Analyzing and modeling real-world phenomena with complex networks: a survey of applications. Advances in Physics 60(3):329–412

Duggirala et al(2017). Duggirala M, Malik M, Kumar S, Hayatnagarkar HG, Balaraman V (2017) Evolving a grounded approach to behavioral composition. In: Simulation Conference (WSC), 2017 Winter, IEEE, pp 4336–4347

Edmonds(1999). Edmonds B (1999) Syntactic measures of complexity. PhD thesis, University of Manchester

Fujiwara and Toida(1982). Fujiwara H, Toida S (1982) The complexity of fault detection: An approach to design for testability. FTCS-12 pp 101–108

Gamma(1995). Gamma E (1995) Design patterns: elements of reusable object-oriented software. Pearson Education India

Gandomi and Haider(2015). Gandomi A, Haider M (2015) Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35(2):137–144

Gavina et al(2018). Gavina MKA, Tahara T, Tainaka Ki, Ito H, Morita S, Ichinose G, Okabe T, Togashi T, Nagatani T, Yoshimura J (2018) Multi-species coexistence in lotka-volterra competitive systems with crowding effects. Scientific reports 8(1):1198

Hayes(2013). Hayes AF (2013) Model templates for process for spss and sas. Retrieved December 12:2013

He and Zhang(2011). He M, Zhang J (2011) A dependency graph approach for fault detection and localization towards secure smart grid. IEEE Transactions on Smart Grid 2(2):342–351

Kavulya et al(2012). Kavulya SP, Joshi K, Di Giandomenico F, Narasimhan P (2012) Failure diagnosis of complex systems. In: Resilience assessment and evaluation of computing systems, Springer, pp 239–261

Kruchten(2012). Kruchten P (2012) Complexity made simple. Proceedings of the Canadian Engineering Education Association

Kumar et al(2017). Kumar S, Duggirala M, Hayatnagarkar H, Balaraman V (2017) Understanding impact of supervisory support on work outcomes using agent based simulation. In: DIAS/EDUDM@ISEC

Laney(2001). Laney D (2001) 3d data management: Controlling data volume, velocity and variety. META Group Research Note 6(70)

Lee(2008). Lee EA (2008) Cyber physical systems: Design challenges. In: Object oriented real-time distributed computing (isorc), 2008 11th ieee international symposium on, IEEE, pp 363–369

Lelkes(2017). Lelkes AD (2017) Algorithms and complexity results for learning and big data. PhD thesis, University of Illinois at Chicago

Lukoianova and Rubin(2014). Lukoianova T, Rubin VL (2014) Veracity roadmap: Is big data objective, truthful and credible? Advances In Classification Research Online

Ming and Vitányi(1997). Ming L, Vitányi P (1997) An introduction to Kolmogorov complexity and its applications. Springer Heidelberg

Newman(2003). Newman ME (2003) The structure and function of complex networks. SIAM review 45(2):167–256

O'Malley(2008). O'Malley O (2008) Terabyte sort on apache hadoop. Yahoo, available online at: http://sortbenchmark org/Yahoo-Hadoop pdf,(May) pp 1–3

Page(2010). Page SE (2010) Diversity and complexity. Princeton University Press

Scattolini(2009). Scattolini R (2009) Architectures for distributed and hierarchical model predictive control–a review. Journal of process control 19(5):723–731

Taylor et al(2013). Taylor SJ, Khan A, Morse KL, Tolk A, Yilmaz L, Zander J (2013) Grand challenges on the theory of modeling and simulation. In: Proceedings of the Symposium on Theory of Modeling & Simulation-DEVS Integrative M&S Symposium, Society for Computer Simulation International, p 34

Vaughn et al(2013). Vaughn MG, DeLisi M, Matto HC (2013) Human behavior: A cell to society approach. John Wiley & Sons

Wand and Friedman(1988). Wand M, Friedman DP (1988) The mystery of the tower revealed: A nonreflective description of the reflective tower. Lisp and Symbolic Computation 1(1):11–38

Ward and Barker(2013). Ward JS, Barker A (2013) Undefined by data: a survey of big data definitions. arXiv preprint arXiv:13095821

Yang(2006). Yang TC (2006) Networked control system: a brief survey. IEE Proceedings-Control Theory and Applications 153(4):403–412

Zaharia et al(2016). Zaharia M, Xin RS, Wendell P, Das T, Armbrust M, Dave A, Meng X, Rosen J, Venkataraman S, Franklin MJ, et al (2016) Apache spark: a unified engine for big data processing. Communications of the ACM 59(11):56–65

Rise of the Neostrategist:
A New Paradigm for the Age of Complexity

Noah J. Komnick
Lieutenant Colonel, USMC

This paper is submitted to the International Conference on Complex Systems (ICCS), held July 22-27, 2018, in Cambridge, MA, USA. It is adapted from an analytic research paper submitted to the Faculty of the United States Naval War College Newport, RI in partial satisfaction for a Master of Arts degree in National Security and Strategic Studies.

May 3, 2018

Abstract

A general theory of strategy (i.e., abstract strategy) has endured since the 18th century. It posits that strategy is the reconciliation of *ends* and *means* in order to determine the *ways*. Inductively, this paradigm was consistently proven valid and strong for over two hundred years, and thus labelled "enduring," by luminary military strategists from Carl von Clausewitz to Colin S. Gray. Yet, as this paper will prove, strategy fundamentally assumes a system is deterministic and thus fails to properly account for the ramifications of complexity. As a result, a new paradigm is proposed: *neostrategy*. Just like the observation of one black swan proves that not all swans are white, neostrategy highlights that strategy is not always useful nor is it enduring. Borrowing from the works of Kenneth O. Stanley, Joel Lehman, and Yaneer Bar-Yam, neostrategy offers planners an alternative to the traditional, objective-seeking strategy that we are so familiar with and instead proposes a *strategy of novelty* for some cases. In the process, this paper explains that some objectives, such as organizational innovation, are intrinsically uncertain and thus better served by a strategy of novelty instead of objective. Therefore, in a Thomas S. Kuhn sense, neostrategy replaces strategy as a more complete paradigm for planning in today's interconnected, globalized, and thus increasingly complex world.

About the Author

Noah J. Komnick is an active duty Marine Corps officer with over 20 years of continued service. His academic degrees include civil engineering (BS), business (MBA), and operational studies (MA). He is also a former student of New England Complex Systems Institute (NECSI) and Santa Fe Institute (SFI). His ongoing research is focused on using complex systems science to solve contemporary problems in national security, strategy, operations, and leadership. His research represents his personal views and is not necessarily endorsed by the United States Marine Corps, Department of Defense, or U.S. Government.

Introduction

The Department of Defense challenges its military officers to ponder two significant questions: how do we change if we do not know the future, and how do we bend the technology curve? These two questions strike at the heart of our contemporary challenge in military strategy—the world is increasingly complex, and technology and its societal impacts are accelerating exponentially, so how can we successfully confront those realities? Will our current understanding and application of military strategy be sufficient to develop tomorrow's operational plans and win future wars? Or, will something new be needed to comprehend, plan, and execute missions in a world increasingly ruled by complexity? To answer these questions, this paper will summarily recapitulate contemporary thoughts on strategy and apply that understanding within a context of complexity which dominates the world today. As a result, a flaw in the logic of strategy is discovered and a new idea, the *neostrategist*, emerges. Whereas the strategist believes the logic of strategy is *unlimited* in application, the *neostrategist* views strategy as inherently *limited* when addressing complex systems. Additionally, the *neostrategist* is aware of the ramifications of complexity and is therefore more apt to develop plans, design forces, and direct operations in scenarios of high complexity. As such, the United States military needs *neostrategists* to solve its security challenges in this new age of complexity.

The structure of this paper's argument is intentionally deductive. First, abstract strategy is summarized as fundamentally unchanged regardless of time, space, and force variables. However, its intrinsic assumption is that the world is deterministic (i.e., with the right input the desired output *will* be achieved). The second premise is that complexity rules the world (social, political, and economic), and its ramifications contradict linear intuition. Said differently, "Isaac Newton was not wrong, his theories are just dangerously incomplete for today's interconnected, globalized world," (Komnick 2017). Therefore, if strategy is immutable yet fundamentally deterministic in nature, and the operational environment is marred in complexity and its associated nonlinearities, then strategy is inherently limited. That said, the first question to answer is: what is strategy?

What is Strategy?

"Strategy has a permanent nature, but an ever-changing character."
– Colin S. Gray, Modern Strategy

The word *strategy* can be interpreted differently, so it is worthwhile to clarify its meaning up front. As Peter Paret explains in *Makers of Modern Strategy: from Machiavelli to the Nuclear Age*, the understanding of strategy has taken many forms throughout history, and its variations are the result of application at different scales and different contexts (Paret 1986). For example, in *On War*, Carl von Clausewitz defines strategy as "the use of engagements for the object of war," (Clausewitz 1989, 128). In today's military vernacular, Clausewitz's strategy is often called operational planning or operational art, where tactical engagements cognitively form a cohesive whole to achieve a military's operational objective. Prior to the 20th century, strategy tended to be confined to the realm of military force in war; it was about how, when, and where to use military force to achieve a military aim and/or political objective. However, during the Interwar period, this long held paradigm about strategy evolved to include the distinctions of not only military strategy but also national strategy. Numerous incarnations of strategy ensued: grand strategy, total strategy, overall strategy, maritime strategy, nuclear strategy, and others (Beaufre 1965). The erudite strategist and military scholar, Colin S. Gray, settles the confusion over strategy by clearly delineating between "strategy and strategies," (Gray 2015). As he explains, strategy is immutable and timeless; it has always persisted as a byproduct of human factors and their perpetual need for politics and policy (i.e., his logic is: humans ⇒ politics ⇒ policy ⇒ strategy).[1] In this sense, he is referring to strategy in the abstract, without context, or as he says, "a general theory of strategy," (Gray 2015, 47). In contrast, while strategy is universal in human affairs, *strategies* are infinitely variable as a result of contextual uniqueness (or, an "ever-changing character," as Gray explains). Gray argues this distinction in response to literature that confuses the two by asserting that because of contextual and/or temporal change strategy must therefore concomitantly change. Gray continues his defense of strategy by arguing against even those who believe strategy may be an illusion (Betts 2000), explaining that the ineffectiveness of *a strategy* does not mean that it is absent. Hence, this is the crucial point to understand about strategy: it always exists if there is a *choice* of action *for a purpose*. More colloquially, if one thinks before taking action there was a strategy. Yet, the particulars of *a strategy* are dependent on its context of variable constituents: *ends*, *ways*, and *means*.

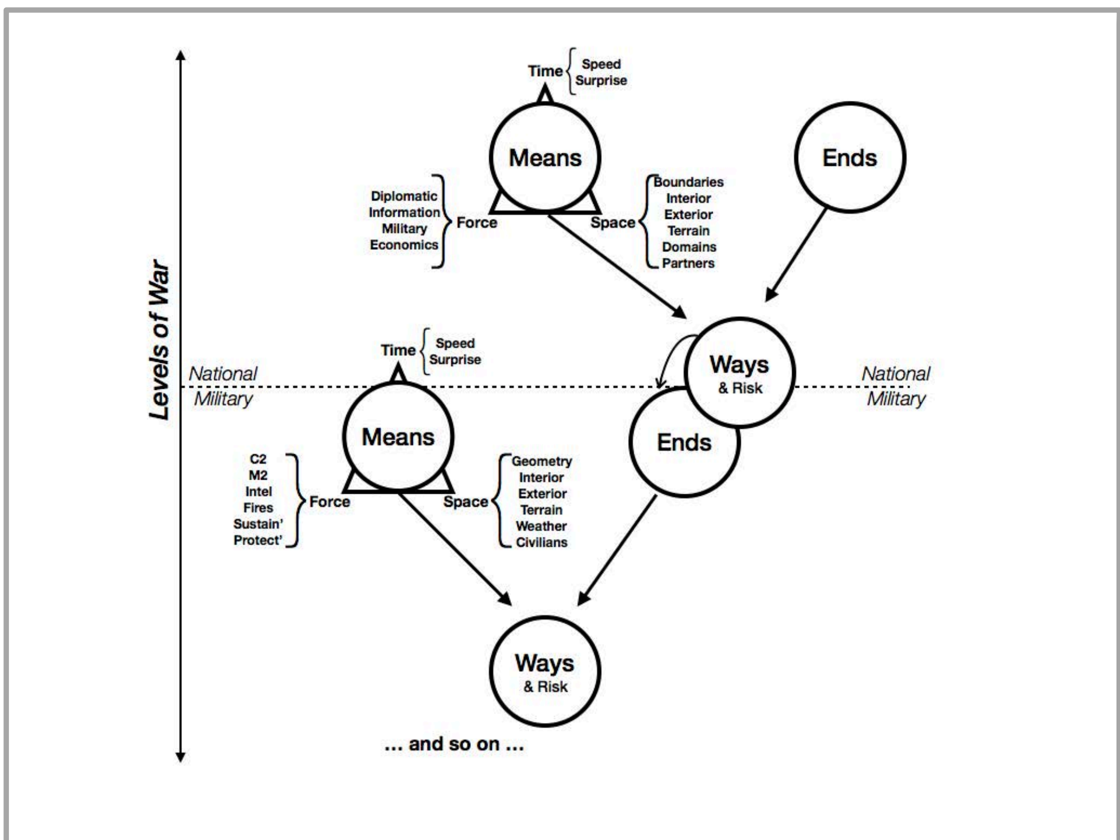


Figure 1: Author's graphical description of strategy at the national and military scales.

Strategy is the decision of how *means* will be used to achieve desired *ends.* Regardless of context, level of war, or warfighting domain, strategy is the reconciliation of

[1] Politics and policy are the process and output, respectively, of deciding who gets what.

ends with *means* in order to determine the *ways* (the how).[2] Figure 1 provides a mental model of this heuristic. The strategist starts with an analysis of the desired *ends*: what is to be achieved in the future or what future event is to be controlled? Next, the strategist reconciles those desired *ends* with the perceived *means* to accomplish those *ends* in order to determine the *ways*. As shown in the figure, the *means* are much more than merely the resources available to the strategist to employ. The *means* include force, time, and space in both absolute and relativistic manners. More colloquially, the *means* are what the strategist has to work with and in, and the *ways* are how s/he will use or cope with the *means* in order to achieve the desired *ends*. Finally, it is important to note that the strategy development process cascades through an organization linking *ways* to *ends* in an apparent hierarchical pattern, but it is more accurately limited by the perceived means of each successive strategist based on his/her position in a collective (e.g., an agency, command, or business). In that sense, the *ways* of one strategist will directly influence the desired *ends* of another (e.g., the *ways* of national strategy become the *ends* of military strategy). In summation, this strategy development process is a sequential heuristic with the intrinsic, and often forgotten, assumption that future events can be controlled if the right *way* is discerned via reconciliation. In other words, strategy is necessarily teleological and it assumes determinism.

Since the idea of strategy has been summarized above, let us now turn to an overview of complexity science and why it challenges our common intuition of linearity and thus strategy itself.

Why Complexity Science?

"Complex problems are the problems that persist—the problems that bounce back and continue to haunt us."
– Yaneer Bar-Yam, President of the New England Complex Systems Institute

Complexity science, the study of complex systems, is relevant to strategy because most problems facing governments today are the byproduct of complex systems. The following paragraphs will explain the origin of complexity science, why it is not yet a customary subject, and why other attempts to apply its lessons to warfighting and strategy have fallen short of making substantial change to how we plan for and fight wars.

Complexity science is a transdisciplinary science comprised of many different and interrelated subjects that together aim to better understand emergent behavior originating from complex systems. Chaos theory, systems theory, complexity theory, nonlinear dynamics, game theory, agent-based modeling, multiscale analysis, and many others fall under the umbrella discipline called complexity science.[3] The origin of complexity science can be traced back to Isaac Newton in the 17th century and Henri Poincaré in the late 19th century. Isaac Newton changed the world with his three laws of motion. It was not merely some feat of physics lore, his laws actually changed how people understood the world to be—from a life ruled by divinity to a world ruled by natural laws. The significant consequence of that revelation was the idea of causality: given initial conditions and natural

---

[2] This author purposely uses the term *reconciliation* to describe the necessary relationship and process between the *ends* and the *means* as it exists in the mind of the strategist. Colin Gray also recognizes this essential relationship, yet he calls it the "strategy bridge."

[3] The Santa Fe Institute defines complexity science as, "the study of emergent system behavior, and seeks to understand how the complex behavior of a whole system arises from its interacting parts." This requires a transdisciplinary approach, borrowing ideas from many different areas (such as the ones listed).

laws, one can now predict future events. The underlying belief was that given enough information (about the positions and velocities of matter) and knowledge of the natural laws, one could predict (and therefore plan for) the future and explain the past. It is in this sense that Newtonian thinking is dominant in human intuition today—understand the rules of X (ecology, biology, government, society, war, etc.) and with enough information one can explain the past and predict the future. However, there was one problem Newton's laws of motion could not solve: the three-body problem. This limitation was far less popularized than his laws or his invention of calculus, and it was not even really understood until Henri Poincaré proved that Newton's three-body problem was indeed unsolvable. Poincaré discovered that because of the interdependencies (dynamics) of the three bodies (in his case planets), he could not accurately predict their orbits because the precision of measurement required to do so was impossible to achieve. The system was so sensitive to initial conditions that even a miniscule error in measurement would cause a dramatic change in the orbits. Today that system characteristic is called *chaos*, an emergent phenomenon often observed in complex systems.[4]

But, what makes a system complex? The answer is that complexity comes from *nonlinear* feedback arising from the interdependencies of a system's parts. As such, the magnitude of complexity is the number of possible states (possibilities) for a given system, and those possibilities are attributable to the system's nonlinear dynamics. *Linear* means that a proportional change of an input equals the same proportional change in output (e.g., double the input leads to a doubling of the output); *nonlinear* means the opposite (e.g., economy of scale—buying more units to spend less per unit—is an example of nonlinearity). Another example of nonlinearity is buying a stock today that exponentially increases or decreases tomorrow due to the interdependencies, thus complexity, of the global equity market. For this reason, some call this type of a system an *interactively complex system* to highlight the dynamic relationship between its individual parts. So, in predicting the behavior of complex systems one needs to understand more than just the system input(s) and its mechanics (e.g., natural laws of motion); dynamic interaction between each of the system's constituent parts must also be accounted for (i.e., system feedback). As Poincaré's and Newton's observations proved, this made forecasting impossibly difficult, and consequently this problem was left unrealized except by a relative few.

It was not until James Gleick's *Chaos: Making a New Science* (1987), that popular interest in complexity science really began. Before Gleick and after Poincaré, many significant breakthroughs in the science came from the likes of Edward Lorenz, Robert May, Mitchell Feigenbaum, and Benoit Mandelbrot, just to name a few. However, their monumental achievements did not receive much recognition outside the physical sciences. Eventually, it was Gleick's book that really caught the attention of the general public and the social sciences, and it is evidenced by how many contemporary books on applying complexity science tend to reference Gleick's book as their starting point for understanding the relatively new science. Naturally, as popular interest grew and the science matured, the literature explaining and applying complexity science to an ever-growing list of topics continued throughout the 1990s and 2000s. The general revelation was this: the world is composed mostly of complex systems, and complex systems require a fundamentally

[4] According to nonlinear dynamics, a system is considered chaotic if it is (1) deterministic, (2) aperiodic (does not repeat), (3) bounded (does not fly off to an infinity), and (4) sensitive to initial conditions.

different understanding from our linear intuition born from traditional (Newtonian) thought. Inevitably, these new ideas would be applied to the subjects of war, strategy, and warfare.

Fusing complexity science with the art and science of war, and its strategy, is not new. For starters, Colonel John Boyd's theories on war and warfare (though unpublished, but nonetheless influential) were likely the first attempt.[5] Subsequent and notable others include: Alan Beyerchen's "Clausewitz, Nonlinearity and the Unpredictability of War," in *International Security* (1992); Roger Beaumont's *War, Chaos, and History* (1994); Thomas J. Czerwinski's *Coping with the* Bounds (1998); Colin S. Gray's *Strategy for Chaos* (2002); Everett Carl Dolman's *Pure Strategy* (2005); Sean T. Lawson's *Nonlinear Science and Warfare* (2014); and, General Stanley McChrystal's *Team of Teams* (2015). Taken together (including others not mentioned here), these treatises into complexity science, war, warfare, and strategy equate to an ongoing dialogue and discovery of practicality of the subject. Yet, each has its own shortcomings, and therefore is why complexity science is still not wholly accepted in contemporary military doctrine or practice. As a result, complexity science is often portrayed as too esoteric and too ambiguous; it is not revolutionary; and, it does not offer anything new beyond the Clausewitzian interpretation of war. Even Colin Gray concludes: "there is nothing really new about this," (Gray 2002, 109). But, he is wrong.

Strategist versus Neostrategist

"The first principle is that you must not fool yourself—and you are the easiest person to fool."
– Richard Feynman

We now arrive at the crux of this paper. Having articulated what strategy is and summarizing the basics and prevalence of complexity, this section introduces and proves that a new way of thinking about strategy is required. This new paradigm is called *neostrategy*. Simply put, strategy is complexity naïve whereas neostrategy is complexity aware. Indeed, there are several prominent lessons of complexity science that pertain and will subsequently evolve one's critical thinking (which is the essence of the *neostrategist*).[6] However, the goal of this paper is not to enunciate all the lessons of complexity science that distinguish a strategist from a neostrategist; that is beyond the scope here. Instead, this paper takes the first, essential step of carefully proving that strategy is not unlimited, it does not work in all situations, and this fact is incompatible with how strategy is understood and practiced today. As a result, the new paradigm of neostrategy will eventually replace the old paradigm of strategy in accordance with Thomas Kuhn's theory of scientific revolutions.

Before addressing the reasons as to why strategy is no longer adequate for coping with complex systems (social, economic, political, etc.), let us first examine the inherent weakness of the argument for strategy: inductive reasoning. To understand the inherent weakness of inductive reasoning, begin by considering the theory that all swans are white. How much information would one need to prove that to be true? One would have to observe all swans everywhere until time infinity for it to be absolutely true. But, as soon as just one black swan is observed, the all-white theory is proven entirely false. Colin S. Gray, our archetypal strategist, makes the same type of argument (inductively) with his general theory of strategy. Using countless historical cases as evidence, Gray and many other strategists

[5] More specifically: Boyd's *Destruction and Creation* (1976), *Patterns of Conflict* (1986), and *Organic Design for Command and Control* (1987). Publicly available here: http://dnipogo.org/john-r-boyd/.
[6] Such lessons include: ergodicity, fragility, power-laws, universality, emergence, and more.

have argued that the general theory of strategy is enduring (i.e., unlimited). Yet, if only one instance is found where the theory does not hold, the theory that strategy is unlimited becomes invalid. That one instance is strategy in a complex system.

The structural reason why the general theory of strategy fails in the context of complex systems is because of the theory's underlying assumption behind the idea of *ends* (goals, objectives, etc.). More precisely, that assumption is called determinism, or assuming that the process of strategy is deterministic (i.e., believing that if only the right strategy is applied it *will* achieve the desired ends). However, when the *means* (the second part of Gray's strategy triptych of ends, ways, and means) are complex then determinism is not guaranteed. Even Gray warns that "faulty assumptions are the most deadly [sic] source of strategic error," and in this case it is the assumption of determinism that is faulty (Gray 2015, 14). But, why is it faulty?

Determinism often does not hold in complex systems because of four reasons: chaos, path dependence, entropy, and cognitive complexity. First, think of strategy development as a search algorithm used to find a solution. In this respect, the strategist searches for a solution *only* in reference to an objective (the desired ends). After all, Gray articulates that "*ends* are the purpose of the endeavor," (Gray 2015, 109). But, this is problematic because of the ramifications of chaotic systems discussed early in this paper. It is true that not all complex systems are chaotic, but when they are chaotic the amount and precision of information needed for a successful strategy (the solution) is impossible (in physics, this is referred to as "chasing Laplace's demon").[7] Secondly, there is the problem of path dependence.[8] Path dependence highlights the fact that finding the right solution is not enough. The strategist must not only find the right solution, but s/he must also determine the right sequencing in space and time. Indeed, sequencing is a regular characteristic in strategies but in complex systems sequencing becomes impossibly difficult to discern due to the interdependencies and nonlinearity, as described earlier, of complex systems. Thirdly, there is the problem of entropy. In this context, entropy is hidden information, which is to say that the unknowable grows with increased complexity (more system possibilities). The more complex a system is, the greater the probability that a solution is hidden and cannot be discovered (i.e., unknown unknowns). The fourth reason is cognitive complexity. Cognitive complexity, whether for an individual or a group, is the ability to acknowledge and evaluate options (possibilities) in isolation from one another (i.e., without bias or interference). The only way to ensure that a successful strategy (a solution) is found in the set of possibilities for a given complex system is if the cognitive complexity of the strategist(s) is/are equal to or greater than the complexity of the system. This principle is derived from W. Ross Ashby's Law of Requisite Variety,[9] and is one reason why Yaneer Bar-Yam, a complex systems scientist and President of the New England Complex Systems Institute, says, "distinguishing realistic goals from fantasies is frequently impossible in a complex system."[10] Said differently, "as complexity increases there are many more wrong ways for every right way to

---

[7] Named after the French scientist, Pierre-Simon Laplace.

[8] In physics, path dependence is called hysteresis, and it refers to future states (e.g., actions, events, decisions) that are dependent on the past sequence of states (in space and time).

[9] The Law of Requisite Variety posits that the degree of control (or understanding) of a system is equal to the ratio of variety between the controller and the given system.

[10] Yaneer Bar-Yam, *@yaneerbaryam*, February 8, 2018.

do something."[11] All this begs the question: can a system be so complex that it becomes impossible to reach the end, goal, or objective?

According to Kenneth Stanley and Joel Lehman, the answer is yes. In their book, *Why Greatness Cannot Be Planned: The Myth of the Objective*, they prove that "ambitious" objectives are *less likely* to occur if they are planned for (Stanley and Lehman 2015, 93). The simplified reason for this is that in complex systems a specific objective is one of nearly an infinite set of possible states for the given system, and the path to that objective is nearly impossible to discern. For example, the goal of flying has existed for centuries, likely longer, yet no one ever planned or foresaw that the combustible engine would have to be invented first. Similarly, vacuum tubes were not invented so man could build a computer in the future; there was no way to foresee that necessary "stepping stone," as Stanley and Lehman call it, to build the first computer (Stanley and Lehman 2015, 29). Still, Stanley and Lehman found that this challenge applies to only select cases—ones in which the objective is intrinsically uncertain. In these cases, the better method is to adopt a strategy of novelty instead of focusing on an objective.

Recall that the essence of strategy is the ends, of ends-ways-means. Stanley's and Lehman's research tells us to replace ends with novelty—the purpose becomes a search for novelty, not an objective or a desired end state. To be clear, novelty is not another variation of objective either. This is because seeking novelty is perpetually elusive; once achieved it is no longer novel. The principle problem with focusing on achieving an objective is that it yields more deceptive stepping stones than searching for novelty, because assessment is based on proximity to or bias towards the objective (consider a maze of transparent walls; it is counterintuitive to walk *away* from the objective even though that might be what is required to succeed).[12]

As mentioned earlier, searching for novelty only applies with objectives that are intrinsically uncertain (i.e., the idea itself is ambitious). Strategies for innovation are likely the most germane. Examples: what is the optimal arrangement and configuration of unmanned aerial vehicles (UAV) with the Marine rifle platoon, or how can artificial intelligence (AI) be used to decrease collateral damage without sacrificing speed and lethality on the battlefield? Right now, the tendency is to solve both these problems by implementing an institutional strategy, led by a centralized command in a given armed service. That effort will likely fail. Conversely, a search for novelty will push *authority to experiment* down to the tactical units *without direction*. As a result, tactical units will experiment at will, trying novel techniques and equipment to enhance combat power, sharing lessons learned (including failures), in order to discover the next evolutionary concept, formation, or weapon system. In the end, to continue the hypothetical, neither UAVs at platoon level or AI on the battlefield may prove useful or cost effective. Instead, what might be discovered is the mobility advantage of bicycles in urban combat (or some other completely *unforeseen* discovery). The point is, searching for novelty is a treasure hunt to find the new innovation that actually leads to greatness. It also has the added benefit that when mistakes or failures occur (as they most surely will), the effects are localized and not systemic (as it would be if the institution fielded new UAVs to all platoons only to find out they do not work as planned; now all platoons are negatively affected [i.e., systemic failure]). Said differently, searching

[11] Yaneer Bar-Yam, *@yaneerbaryam*, November 19, 2017.
[12] Stanley and Lehman admit that the merits of searching for novelty may be tough to accept because it is counterintuitive, but their book goes into great detail in rebutting the criticisms and skepticism of the idea.

for novelty works much like biological evolution: genetic mutation ⇒ survive or die ⇒ reproduce (traits that lead to increased survival are copied via reproduction). So, as Stanley and Lehman assert, the path to greatness is better done by searching for novelty.

Finally, let us return to this paper's main argument that strategy is limited and concomitantly a new paradigm of neostrategy is warranted. The efficacy of searching for novelty, in addition to the four reasons explained earlier, proves that not all systems are deterministic (the fundamental assumption of strategy). If determinism fails in only one case, then the logic of strategy cannot be "permanent in nature."[13] To be colloquial, sometimes you really cannot get there from here. Now, one could argue that in those cases strategy still matters even though it did not work. However, the fact that it will not work is known *before* execution, not after; and, if a strategy is known not to work at the time of its inception, then it is not a strategy (otherwise anything can be a strategy). Hence, a neostrategist realizes this limitation of strategy and is therefore better able to cope with the realities of an increasingly complex world.

Conclusion

"Not everything is possible," (Stanley and Lehman 2015, 55). That is the crucial point. Strategy, as a logic, does not account for that reality, and it is therefore limited and not enduring in nature. For that reason, contemporary strategists tend to believe that anything *is* possible, and often quip, "we just need to find the right strategy." That is Newtonian intuition talking, and possibly American optimism too. In truth, sometimes you really "cannot get there from here," as the old New England saying goes. In an age of increasing complexity, where there are more wrong answers for every right one, the *neostrategist* will understand a new way, searching for novelty, is sometimes necessary. Moreover, with the accelerating pace of changing technologies and the associated changes in societies, political systems, and economic markets, top-down driven strategies will likely fail even more often than they do today. The trademark of the neostrategist is understanding these limits of control and information in complex systems. As such, the neostrategist would answer the Department of Defense's leading question, "how do we bend the technology [or complexity] curve," with: "We don't; we *ride* the curve by experimenting at the lowest tactical level possible."

[13] From earlier Colin Gray quote on page 1.

Bibliography

Bar-Yam, Yaneer. 2004. *Making Things Work: Solving Complex Problems in a Complex World.* Cambridge, MA: Knowledge Press.

Beaufre, André. 1965. *Introduction to Strategy.* London: Faber & Faber.

Beaumont, Roger. 1994. *War, Chaos, and History.* Westport, CT: Praeger.

Bousquet, Antoine. 2009. *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity.* New York: Columbia University Press.

Clausewitz, Carl von. 1989. *On War.* Edited by Michael Howard and Peter Paret. Princeton, NJ: Princeton University Press.

Dolman, Everett Carl. 2005. *Pure Strategy: Power and Principle in the Space and Information Age.* New York: Frank Cass.

Du Sautoy, Marcus. 2016. *The Great Unknown: Seven Journeys to the Frontiers of Science.* New York: Viking.

Gray, Colin S. 2002. *Strategy For Chaos: Revolutions in Military Affairs.* London: Frank Cass.

—. 2015. *The Future of Strategy.* Malden, MA: Pality Press.

Gribbin, John. 2004. *Deep Simplicity: Bringing Order to Chaos and Complexity.* New York: Random House, Inc.

Komnick, Noah J. 2017. *MWCS-38: What We Have Done, What We Have Learned.* Post-command Brief, San Deigo: unpublished.

Lawson, Sean T. 2014. *Nonlinear Science and Warfare: Chaos, Complexity and the U.S. Military in the Information Age.* New York: Routledge.

Paret, Peter. 1986. *Makers of Modern Strategy: from Machiavelli to the Nuclear Age.* Princeton, NJ: Princeton University Press.

Stanley, Kenneth O., and Joel Lehman. 2015. *Why Greatness Cannot Be Planned: The Myth of the Objective.* New York: Springer.

Strogatz, Steven H. 2015. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering.* Boulder, CO: Westview Press.

West, Geoffrey. 2017. *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Like in Organisms, Cities, Economies, and Companies.* New York: Penguin Press.

ARTIFICIAL INTELLIGENCE AND LEGAL PERSONALITY: ANY RESCUE FROM *SALOMON V. SALOMON*?

Damilola O. Odumosu[1] and Solomon T. Grace[2];

1. Segun Sipeolu & Associates; Ikeja, Lagos, Nigeria. damiodus@hotmail.com
2. Matrix-Solicitors LLP; Ikoyi, Lagos, Nigeria. gracetosinyourcounsel@gmail.com

ABSTRACT

Most scientific conferences tend to ignore the contributions of law to the development of Artificial Intelligence (AI). This paper tends to bridge this gap by considering the multidisciplinary perspective of Artificial Intelligence as a developing discipline and the legal quandary it as thrown the law Courts. Considering, the advancements attained by Artificial Intelligence so far, should the Courts upon trial find the manufacturers liable for defects of its autonomous device or should a robot simply be confiscated and destroyed? Our legislators, policy makers, academicians, legal scholars and judges should be properly guided by providing from a laboratory of refined thoughts and practical ideas, a definite approach towards ensuring that a robot which cannot be sued today, will not only be able to be sued tomorrow but also be trained on causing minimal or no damage. As one Court observed, "robots cannot be sued," even though they can cause devastating damage." The Courts from time immemorial have always forged new path by ensuring that the principle of '*ibi jus ibi remedium'* applied in circumstances that presents itself just and fair. It is widely believed that within the ambit of every Court there lies the power to find a defaulting party liable. In 2010, it was reported that a robot made by a Swiss art group purchased arms from a black-market website and was later arrested by the Italian police who could not prosecute further because no law recognized such. This has increased the level of knowledge on whether Artificial Intelligence should be considered as a legal person. In considering Artificial Intelligence as a person, would a piece of legislation solve this question or pronouncements from a court of law solve the legal puzzle? Can we also take a bold step by understanding why exactly Artificial Intelligence should be considered as an artificial person by examining how it leans and adapts to the human society.

Keywords: Artificial Intelligence, Robots and Law.

1 INTRODUCTION

It is a great pleasure and an extension of our sincere gratitude to the organizers of this conference as well as the entire management team of the New England Complex Systems Institute for the opportunity to address the distinguished and scholarly scientific community at this 9th International Conference on Complex Systems. This conference is very historic and synonymous in relation to its theme and location with the first International Conference on Artificial Intelligence and Law which was held here in Massachusetts, Boston from 27th to 29th May, 1987. Permit me to introduce to you a very recent discussion and a reflection of our thoughts on the legal rights of Artificial Intelligence (hereinafter, "AI") with this topic titled, 'Artificial Intelligence and Legal Personality: Any rescue from *Salomon V. Salomon*?'

The word artificial intelligence was coined by John McCarthy in a summer workshop held in 1956 Dartmouth College. There is a dearth of a globally acceptable definition of AI both in the academic as well as within the legal community. AI refers to computer systems that think and act like humans, and think and act rationally. [1] Merriam Webster defined artificial intelligence as a branch of computer science dealing with the simulation of intelligent behavior in computers. [2] More so, the Encyclopedia Britannica defines "artificial intelligence, as the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings." Intelligent beings are those that can adapt to changing circumstances. [2]

Advances in AI have shown the greater potentials of AI being better than humans. For instance, autonomous machines can execute complex financial transactions, flag potential terrorists using facial recognition software and also perform document review. [3] The performances of AI within the medical profession are phenomenal. As the human body was not designed to be a machine factory or to work unlimitedly [4] the need to rely on AI to help make human lives better cannot be underestimated. The increasing potency of the usefulness of Artificial Intelligence (AI) systems are found in the recent feats and achievements it as attained. With each passing month, AI as gained footholds in new industries and becomes more enmeshed in our day-to-day lives, and that trend seems likely to continue for the foreseeable future. [5] The potential for further rapid advances in AI has prompted utterances of alarm signifying fears from many quarters, including some calls for government regulation of AI development and restrictions on AI operation. [6]

In Nigeria, the rays of the invasion of Artificial Intelligence are now visible within the socio economic strata and not yet recognized within the legal community in relation to service delivery to clients and in Courts. [7] No doubts exist such that AI as spread its tentacles towards Africa that amidst the high rate of unemployment, there is little or nothing that can be done to stop it. A recent prediction posits that automation will displace nearly 13 percent of South Africa's current work activities by 2020; [8] Robot workers have diminished the bargaining powers of the labour unions representing cashiers and shop assistants. [9]

It is rather sad that there is a dearth of legislation on Artificial Intelligence, automation of drones, robots or AI systems in Nigeria. The Nigerian economy has a lot of artisan vocation that will be swept by the tsunami of AI's potency. Jim Yong, the World Bank President, in a recent assertion predicts that automation will wipe out two-thirds of jobs in the developing countries, with about 50 per cent of jobs across various industries now taken over by robots. [9] This is already evolving.

Modern human history records successful landing on the moon by man, installation of functional space station and so many other space mission success stories. The greatest feat of robots (Spirit and Opportunity) was not only their landing on Mars, but a successful exploration on that planet. [10] This is an achievement that man may not yet accomplish. It must be quickly noted however, that AI will definitely encounter its limitations in Africa in terms of regular power supply, expertise and professionalism.

This first part of this paper states briefly the doctrine of legal personality as enunciated by Lord Halsbury in the case of ***SALOMON V. SALOMON CO. LTD*** [11]. The legal recognition of artificial intelligence using its relevance in the legal profession as well as recent developments in the field of AI is examined in Part II. Part III ushers in the Nigerian perspective to AI in respect of Nigerian laws with a focus on the Nigerian constitution as well as the Intellectual Property regulations and other laws on technology in Nigeria. The final part concludes.

PART I

2 THE DOCTRINE OF LEGAL PERSONALITY

The doctrine of legal personality is a universal legal concept as seen in the case of ***SALOMON V. SALOMON*** **(SUPRA)** by virtue of the landmark judgment delivered by Lord Halsbury which birthed the doctrine of legal personality that separated the shareholders, promoters, directors, and management from the company upon its incorporation. Simply put, upon incorporation, a company acquires a new, separate and distinct personality in law thereby conferring upon it the powers, duties and rights that a natural person enjoys. This distinct and separate personality is often referred to as the veil of incorporation. This doctrine is illustrated in Fig. 1, below which explains the pre incorporation as well as the post incorporation of a company.

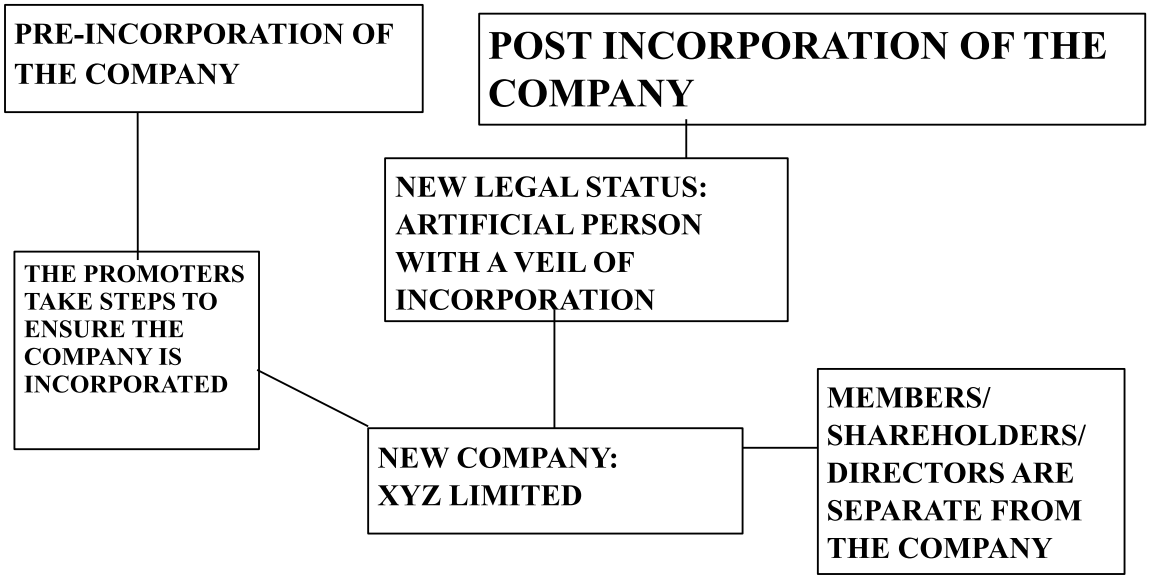


Fig 1: Doctrine of Legal Personality: Pre-Incorporation and Post Incorporation

The Companies and Allied Matters Act (hereinafter, CAMA) is the principal law that regulates the formation, registration, incorporation and winding up of a company in Nigeria. By virtue of the relevant provision of the CAMA. [12] A company from the date of its incorporation becomes a body corporate capable of exercising all the powers and functions of an incorporated company including power to hold land and having perpetual succession and a common seal.

The Apex Court in Nigeria has affirmed this position on legal personality in the case of ***EMENITE LTD. V. CHIEF COLUMBUS E. OLEKA*** [13] were it held that 'the legal personality of a corporate body can only be established as a matter of law by the production in evidence of the certificate of incorporation, but that where, by the state of pleading, the legal personality of a corporate body is not in issue, there will be no need to prove the status and legal personality of a corporate body. There are advantages embedded in incorporation of a company such as, perpetual succession, right to own property – the company can purchase and sell properties in its own name, liabilities of the members are different from the company – upon incorporation the liability of the company becomes separate from its members, legal capacity i.e. to sue and be sued – the company can sue a third party and vice versa, borrowing powers – a company can borrow from a financial institution using its assets as a security, liable in contracts and torts – the can enter into a contractual relationship with a third party and may be liable for any tortious damage, negligence or nuisance, formation of another company – by forming a new company it creates a chain/group of companies whereby the former is a holding company while the latter is the subsidiary company, and transfer of shares – shares can be transferred by the company to a third party, another shareholder or sold outrightly.' [14]

PART II

3 THOUGHTS ON LEGAL RIGHTS OF ARTIFICIAL INTELLIGENCE

Legal rights of AI will enable it to exercise rights, duties and functions which humans are already enjoying both at the national and the international levels. This recognition will create both civil and criminal liabilities as explained by the illustration Fig. 2. There may also be need to regulate AI at both the domestic and international planes which will call for questions relating to human rights, privacy, intellectual property, alienation rights (real estate) while regulating AI as a whole. The question relating to who should be held criminally responsible for genocide, war crimes or crimes against humanity committed by (fully or partially) autonomous machines enabled Artificial

Intelligence like drones or robots will be solved partially. [15] The enabling legal provision will facilitate the powers of the Courts in exercising jurisdiction over AI systems, fully autonomous robots, drones, semi or partially controlled robots in both civil and criminal matters. [15] The police and other law enforcement agencies of government with prosecutorial powers will find common grounds for arresting, detaining, investigating and prosecuting a robot.

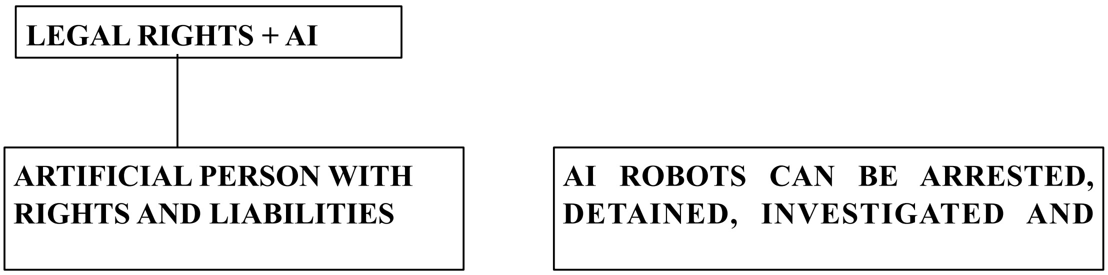


Fig 2: Legal Rights of Artificial Intelligence changes its status to an Artificial Person.

The second reasoning in this regard estimates the level of reliance of humans on AI in the fields led by professionals, highly trained personnel; government certified persons, such that the latter now rely on the former for advanced technical information produced at lightning speed which will require days in human time to process and access. The comparison between the most active branches of AI (*Robotics, Planning, Speech, Vision, Expert Systems, Natural Language Processing* (NLP) and *Machine Learning* (ML) [16] and humans, evidences the enormous capacity of the former to learn, adapt and develop faster than the best of humans combined over time. [16]

In 2011, National Aeronautics and Space Administration (NASA), developed Robonaut – a state-of-the-art, highly dexterous, humanoid robot: Robonaut 2 (R2). R2 is made up of multiple component technologies and systems, vision systems, image recognition systems, sensor integrations, tendon hands, control algorithms, and much more. [17] This humanoid robot buttresses and stresses to a great extent, the second reason adduced in support of the legal recognition of artificial person as an artificial person.

The third approach in this respect is that since AI learns with the aid of machine learning and natural language processing granting legal recognition will enable AI to understand and appreciate human values, ethics, norms, laws and disciplinary actions made for humans. Both ML and NLP are applications of AI. Machine learning is an application of AI with the ability for a program to learn certain tasks and improve skills with direction and/or feedback from humans while Natural Language Processing another AI application with the ability to understand verbal or written natural language queries and provide meaningful conversational responses. Since most regulations are written and easily available AI will encounter no difficulties in understanding the purpose of a regulation.

Humans in some of the world's most regulated profession have either been disciplined, cautioned or had their license revoked due to a serious infraction of some of the rules regulating their profession. It is advocated that this disciplinary approach to sanctioning of humans should be applied to AI especially when it has been used for over a period of time; in a circumstance where it takes an indecisive action which resulting into negligence occasioning harm to humans; such an AI upon legal recognition, should be sanctioned. This will enable other AI artificial persons learn from this. In the human community a sanction of one is deterrence to others; Fig.3 illustrates how AI will be disciplined. This assertion draws strength from the three laws of robotics devised by the late science fiction writer, Isaac Asimov:

1. *A robot may not injure a human being or through inaction allow a human being to come to harm.*

2. *A robot must obey the orders given it by human beings, except where such orders would conflict with the first law.*

3. *A robot must protect its own existence as long as such protection does not conflict with the First and Second Law. [18]*

This author agrees entirely with co-authors J. G. Castel and M. E. Castel, wherein both argued that when writing software for autonomous artificial intelligence machines of general intelligence to become super-intelligent, programmers could incorporate the first two Asimov laws as well as the "general principles of law recognized by civilized nations", human ethical principles and moral values. [15]

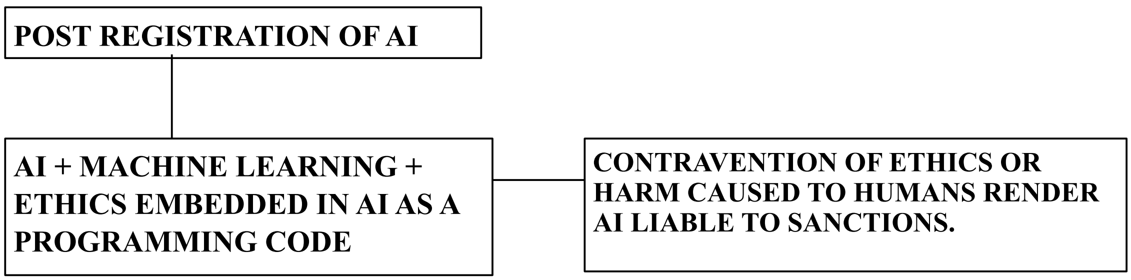


Fig. 3: Discipline of AI with the aid of Machine learning

Recently, Sophia was granted citizenship by the Saudi Arabian Government. Sophia was grateful and she thanked the Kingdom of Saudi Arabia. This made her very honored and proud of the unique distinction. She recognized the fact that it was historic to be the first robot in the world to be recognized with citizenship. [19] Sophia's citizenship has opened for her a legal recognition under the laws of the Saudi Government. She will now enjoy rights of a Saudi Citizen as stated in the Saudi Basic Law. [20]

Artificial Intelligence has been very useful within the legal community such that AI appears in lawyers' workflows everyday – although perhaps not as obviously as in our personal lives. [21] What is helpful for our professional lives is that we are starting to leverage AI across a broad range of applications: legal research, litigation strategy, e-Discovery, self-help online legal services, dispute-resolution models, and contract review and analysis. [22]

In relation to legal research, AI has greatly improved research outcomes. For instance, Westlaw® Answers is an example – it can provide answers to common, well-defined types of legal questions – about statutes of limitation that is, elements of specific causes of action; [23] contains specific answers to common legal questions from authoritative court decisions. Simply type in phrases like "breach of contract" and you will have instant access to a list of cases that lay out core concepts like the elements. [24] With litigation strategy, AI classifies and organizes data faster, better and cheaper, augments human intelligence, empowers people to make use of huge amounts of data to make better decisions and tell better stories. [25]

Like the Latin maxim, ***res ipsa loquitor*,** the overall benefits embedded in AI speaks for itself; it only creates a means for which AI should be considered for regulation. Needless to say that AI is ripe enough for statutory regulation not only in Nigeria but all over the technological world. However, it needs be stressed that regulation would not make AI over-reaching; rather statutes must recognize it first as an artificial person and not just a mere technological trend that requires a robust or overhauling legislation. With the recent enthronement of the General Data Protection Regulation by the European Parliament AI programmers will be required to incorporate some of the provisions contained in the regulation into AI.

As earlier alluded to in this paper the registration of a company in law separates it from its promoters, directors and shareholders. The need to register AI will be necessary upon the legal recognition of the rights of AI as an artificial person thereby enabling the law to monitor the activities of AI in accordance to regulation. It should be borne in mind that the essence of regulating AI is not to stifle its innovative capacity rather to ensure that it does not harm any human at any level. The recent killing of a pedestrian by a driverless vehicle has called for questions on whether AI vehicles or robots will at all be safe within the human community and if manufacturers can be held liable for any damage caused by AI.

Finally, it can be logically argued that upon recognition of AI as an artificial person the advantages embedded in incorporation such as, perpetual succession, right to own property, liability of the members different from the company, legal capacity i.e. to sue and be sued, borrowing powers, liability in contracts and torts, formation of another company and transfer of shares should be fused into any piece of legislation that will regulate the recognition of AI as an artificial person. Unlike a company, AI should be registered for a specific purpose. This will ensure that registered user of AI are held accountable for what it does. Failure to register AI will render the user strictly liable for any harm or injury caused by AI. Fig. 4 illustrates how AI's legal recognition makes it distinct from its owners or users.

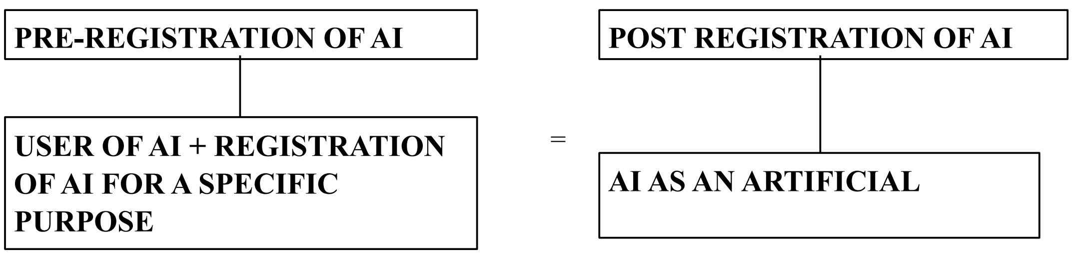


Fig. 4: AI and users: A distinctive feature.

PART III

4 ARTIFICIAL INTELLIGENCE AND THE CONSTITUTION OF NIGERIA, 1999 (AS AMENDED 2011)

Nigeria operates a federal system of government that is regulated by the 1999 constitution (as amended 2011). The Constitution is the Supreme Law in Nigeria. [27] Its Chapter IV, [28] which is enforceable in a Court of law, provides for a number of rights which all Nigerian citizens are entitled to. Such rights include right to life, right to freedom of association, right to freedom of speech, right to own property anywhere within Nigeria to mention but a few.

The Constitution, however, does not recognize artificial intelligence in any manner whatsoever. No mention was ever made of it. Sections 36, 38, 39, 43 and 44 provides for the right to fair hearing, right to freedom of religion, right to freedom of expression, right to own property anywhere in Nigeria and right to adequate compensation for compulsorily acquired land respectively.

It is arguable however, that by the combined provision of relevant sections of the Patents and Designs Act, NOTAP Act and CAMA – Sections 2 and 3 Patents and Designs Act, Section 5 (1) of the National Office for Technology Acquisition and Promotion Act and Section 37 of the Companies and Allied Matters Act, once the robot or machine is registered under the relevant law, it may be accorded a legal personality.

What may not be clear again is whether its recognition should be on the Exclusive Legislative List or Concurrent Legislative List of the 1999 Constitution (as amended in 2011). Suffice to say that since Nigeria practices democracy and the presence of AI will not encroach on the federal-state dichotomy, it should be accorded recognition on the concurrent legislative list. Modern governance recognizes the rights accruing to each federating units in a federation. Its presence requires an enabling law to give it life. Each state of the federation should legislate on it. Without recognition of AI by either the Constitution or any enabling law(s), AI may not enjoy the benefits of being a citizen of Nigeria.

4.1 INTELLECTUAL PROPERTY AND TECHNOLOGY LAWS IN NIGERIA

In Nigeria, our laws on technology have not yet recognised Artificial Intelligence as a whole. However, there are some laws that require parties to register any technology that is to be used in Nigeria. By virtue of these registration AI robots, drones or devices can be used in commercial or industrial parlance. The National Office for Technology Acquisition and Promotion Act (NOTAP) [29] empowers the NOTAP to monitor the transfer of foreign technology to Nigeria.

By the provisions of this Act, it is mandatory to register all contracts for the transfer of foreign technology entered into by any person in Nigeria with NOTAP [30] as well as registration of every contract or agreement entered into by any person in Nigeria with a person outside Nigeria within the time frame of 60 days from the execution or conclusion of the contract. [31]

Contracts are registrable under the Act if its purpose or intent is, in the opinion of NOTAP, wholly or partially in connection with any of the following purpose: the use of trademarks; the right to use patented inventions; the supply of technical expertise in the form of the preparation of plans, diagrams, operating manuals or any other form of technical assistance of any description whatsoever; the supply of basic or detailed engineering; the supply of machinery and plant and the provision of operating staff or managerial assistance and the training of personnel. [32]

Significantly, in Nigeria, technologies and innovations are generally protected by Intellectual Property (IP) Laws. The Copyright Act [33] protects literary work which includes computer programs [34] while the Patents and Designs Act [35] protects industrial designs as well as inventions which are new or an improvement upon an existing patented invention. There is also the Trade Marks Act [36] which protects the owners of registered trademarks. It is worthy of note that, owners of unregistered trademarks are not protected by the Trade Marks Act but are entitled to seek relief under the English Common Law principles which are applicable in Nigeria. An unprotected trademark can be enforced by instituting a legal action for passing off at any of the State High Courts in Nigeria.

The National Information Technology Development Agency Act (NITDA) [37] NITDA was created by an act of parliament in April 2001 to implement the Nigerian Information Technology Policy and co-ordinate general IT development in the country. It was mandated by the 2007 National Information Technology Development Act to create a framework for the planning, research, development, standardization, application, coordination, monitoring, evaluation and regulation of Information Technology practices, activities and systems in Nigeria. Its role therefore is to develop Information Technology in the country through regulatory standards, guidelines and policies.

Additionally, NITDA is the clearing house for all IT projects and infrastructural development in Nigeria. It is the prime Agency for e-government implementation, Internet governance and General IT development in Nigeria. [38] The responsibility of midwifing AI usage in Nigeria rests on this organization. And this can be done through stakeholder conferences and public hearing. After this, the relevant unit of the national assembly can then be consulted for necessary steps to be taken for the law to be enacted.

Finally, there is a recognized technology law society in Nigeria known as Technology Law Society of Nigeria (TLSN). It is a non-profit organization founded in order to encourage the growth of all aspect of technology and electronic transaction law in Nigeria. To that extent, it is saddled with the duty to collaborate with all legal entities and serve as a common meeting point for all electronic transactions, technology law experts, consultants, scholars, other legal professionals, jurists and other stakeholders who had an interest in the evolution, growth and development of electronic transaction and technology law. [39]

5 CONCLUSION

AI is indeed the next big industrial revolution which has the capacity to develop and bridge the intellectual gap, literacy awareness and broaden the minds of humans within a short time. AI's capacity to provide the legal world and humanity at large a requisite and better value for time and efforts spent on research can simply be transited if and only if AI is considered as an artificial person which will definitely provide new opportunities. The Constitution of the Federal Republic of Nigeria, 1999 (as amended in 2011) by its provision is wide enough to accommodate this new phenomenon. It will assist not only in the legal profession but also in other disciplines like medicine, military, the police and host of others. For Nigeria and indeed Africa at large, it will be great to consider the opportunities provided by AI, understanding AI and providing a natural habitat for its thriving potentials could be encouraged by regulating AI's activities which Nigerian citizens especially legal practitioners will greatly value and appreciate.

REFERENCES

1. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach (3rd Edition)* (Essex, England: Pearson, 2009).
2. Bernard Marr, The Key Definitions Of Artificial Intelligence That Explains Its Importance, (Forbes, February 14, 2018) https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#1d80aaad4f5d accessed June 8th, 2018.
3. Timothy Williams, Facial Recognition Software Moves From Overseas Wars to Local Police (New York Times, August 13th 2015) <https://www.nytimes.com/2015/08/13/us/facial-recognition-software-moves-from-overseas-wars-to-local-police.html> accessed March 1st, 2018
4. Chima Akwaja, Tony Awunor, Jide Fabamise And Bukola Idowu, (Leadership, 28 May 2017) <https://leadership.ng/2017/05/28/nigeria-nations-lose-50-jobs-automation/ > accessed March 1st, 2018
5. Kevin Kelly, The Three Breakthroughs That Have Finally Unleashed AI on the World (Wired, October 27th, 2014) <https://www.wired.com/2014/10/future-of-artificial-intelligence/> accessed 1st March, 2018
6. John Frank Weaver, We Need to Pass Legislation on Artificial Intelligence Early Often (Slate, September 12th 2014) www.slate.com/blogs/future_tense/2014/09/12/we_need_to_pass_artificial_intelligence_laws_early_and_often.html accessed March 1st, 2018.
7. As at the year 2018, there's a dearth of legislation on autonomous vehicles, drones and artificial intelligence. More so, Nigerian Law Firms and Courts have not begun using AI for client service delivery, document review or legal representation.
8. Lexi Novitske, the AI Invasion is Coming to Africa (and It's a Good Thing), (Stanford Social Innovative Review, February 12th, 2018) <https://ssir.org/articles/entry/the_ai_invasion_is_coming_to_africa_and_its_a_good_thing> accessed March 4th, 2018.
9. Chima Akwaja, Tony Awunor, Jide Fabamise And Bukola Idowu, (Leadership, 28 May 2017) <https://leadership.ng/2017/05/28/nigeria-nations-lose-50-jobs-automation/> accessed March 1st, 2018
10. International Federation of Robotics, < https://ifr.org/robot-history> accessed March 7th, 2018.
11. [1897] AC 22
12. Companies and Allied Matters Act, Section 37, CAP C20, Laws of the Federation, 2004.
13. (2005) 6 NWLR (Pt. 921) 350, Emenite Ltd v. Chief Columbus E. Oleka (2004) LPELR-CA/E/149M/2001
14. Rotimi Kodaolu, Nigeria Company Law: The Principle of Legal Personality and The Protection of Minority Members and Other Non-Members Constituents, <www.academia.edu/8439040/Nigeria_Company_Law_The_Principle_Of_Legal_Personalityand_The_Protection_Of_Minority_Members_And_Other_Non-Members_Constituents> accessed April 13th, 2018
15. J.-G. Castel and Matthew E. Castel, The Road To Artificial Superintelligence: Has International Law A Role To Play? Volume 14, No. 1, Canadian Journal on Technology and Law, < https://ojs.library.dal.ca/CJLT/article/download/7211/6256 > accessed on March 7th, 2018

16. Artificial Intelligence in Law: The State of Play 2016, Part 1, (Neota Logic, 28 February 2016) <https://www.neotalogic.com/2016/02/28/artificial-intelligence-in-law-the-state-of-play-2016-part-1/> accessed March 13th, 2018. It is no longer news that AI enabled computers are better than humans in fields where they are genius or experts. It's the dramatic outcome of artificial intelligence research at Google's Deep Mind subsidiary, whose AlphaGo program recently won five straight against the top-ranked Go master in Europe. The game of Go is 2,500 years old and, despite its simple rules, is many orders of magnitude more complex than chess. See also, Elizabeth Gibney, Google AI algorithm masters ancient game of Go Deep-learning software defeats human professional for first time, (Nature, January 27th, 2016) <www.nature.com/news/google-ai-algorithm-masters-ancient-game-of-go-1.19234> on the March 13th, 2018
17. National Aeronautics and Space Administration (NASA), https://robonaut.jsc.nasa.gov/R2/ accessed March 25th, 2018
18. Isaac Asimov, "Runaround" (1942) published in *I Robot* (New York, NY): Grome Press, 1950)
19. Andrew Griffin, Saudi Arabia Grants Citizenship to a Robot for the First Time Ever' (Independent, 26 October, 2017) <https://www.independent.co.uk/life-style/gadgets-and-tech/news/saudi-arabia-robot-sophia-citizenship-android-riyadh-citizen-passport-future-a8021601.html> accessed 15th March, 2018.
20. Articles 23 – 43, *Saudi Arabia – Constitution* (Adopted by Royal decree of King Fahd on March 1992) See also <http://www.the-saudi.net/saudi-arabia/saudi-constitution.htm> accessed March 15th, 2018.
21. Thomson Reuters White Paper Series, Legal AI for the Business and Practice of Law, <https://static.legalsolutions.thomsonreuters.com/static/pdf/Demystifying-Artificial-Intelligence.pdf> accessed on March 19th, 2018
22. Ryan Pottebaum, Six legal research tools to know about in 2017, <https://legalsolutions.thomsonreuters.com/law-products/westlaw-legal-research/insights/six-legal-research-tools-to-know-about-in-2017> (Thomson Reuters, 2017) accessed March 19th, 2018
23. Julie Sobowale, How artificial intelligence is transforming the legal profession, <www.abajournal.com/magazine/article/how_artificial_intelligence_is_transforming_the_legal_profession> (ABA Journal, April 2016) accessed March 19th, 2018
24. Where Next for Premonition, the AI-Driven Litigation Analysis Platform? (Artificial Lawyer, 20 February, 2017) <https://www.artificiallawyer.com/2017/02/20/where-next-for-premonition-the-ai-driven-litigation-analysis-platform/> accessed March 19th, 2018.
25. Lex Machina, < https://lexmachina.com/legal-analytics/> accessed March 19th, 2018
26. Jeff Pfeifer, How Analytics Is Shaping the Current and Future Practice of Law, (Law Journal Newsletters, 1 July 2017) <www.lawjournalnewsletters.com/sites/lawjournalnewsletters/2017/07/01/how-analytics-is-shaping-the-current-and-future-practice-of-law-4/?slreturn=20180219130212> accessed 19th March, 2018
27. Section 1 (1) 1999 Constitution (as amended)
28. Sections 33 – 46 1999 Constitution (as amended)
29. National Office for Technology Acquisition and Promotion Act (NOTAP), Cap N62, Laws of the Federation 2004
30. Section 5 (1) NOTAP Act
31. Section 5 (2) NOTAP Act
32. Section 4 d (i –vi) NOTAP Act
33. Copyright Act, CAP C28 LFN 2004
34. See Section 1(2)(b)A Copyright Act; literary musical or artistic work shall not be eligible for copyright unless (b) the work has been fixed in any definite medium of expression now known or later to be developed from which it can be perceived reproduced or otherwise communicated either directly or with the aid of any machine or device. See also Jide Babafemi Ph.D, *Intellectual Property: The Law and Practice of Copyright, Trade Marks, Patents and Industrial Designs in Nigeria*, 1st Ed., Justinian Books Limited, 2007, p. 8. See the Nigerian case of Adenuga v. Ilesanmi Press (1991) 5 N.W.L.R. 82 and the English case of Sweeney & Anor v. MacMillan Publishers Ltd & Anor, Court of Appeal - Chancery Division, December 19th, 2001, [2001] EWHC Ch 460; <http://lexisweb.co.uk/cases/2001/november/sweeney-and-another-v-macmillan-publishers-ltd-and-another> (Lexisweb.co.uk, November 2001) accessed 16th March, 2018.
35. Patents and Designs Act, CAP P2 LFN 2004
36. Trade Marks Act, CAP T13 LFN 2004
37. National Information Technology Development Agency Act, CAP N156 LFN 2004

38. National Information Technology Development Agency, < http://nitda.gov.ng/about-nitda/> accessed March 19th, 2018.
39. Dr. Feyi-Sobanjo, O. T. <http://www.techlawnigeria.org/profile.html> (Technology Law Society of Nigeria) accessed on the March 19th, 2018

Consequences of Changes in Global Patterns of Human Interaction

Anne-Marie Grisogono[1*][0000-0002-4412-2187], Roger Bradbury[2][0000-0002-0940-3493], John Finnigan[2 3][0000-0003-1073-0886], Nicholas Lyall[2][0000-0002-6784-9311], Dmitry Brizhinev[2]

[1]Flinders University, Adelaide, Australia

[2]Australian National University, Canberra, Australia

[3]Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia

[*]annemarie.grisogono@flinders.edu.au

Recent rapid and extensive changes in global patterns of interaction between individuals resulting from exponential growth in the proportion of populations participating in social media and other interactive online applications, suggest a number of possible consequences some of which are concerning for the future of democratic societies and for the stability of global order. We draw insights from the scientific study of collective phenomena in complex systems. Changes in interaction patterns often bring about phase transitions – system rearrangements that are sudden and transformative, through the emergence and self-amplification of large-scale collective behaviour. We see a parallel here in the possible effects of changes in human interaction patterns on the structure of global social systems – including the traditional structures of nation states, and national and cultural identities. Since the growth of new groupings is now largely driven by recommender algorithms in social applications, whereby people become more and more connected to likeminded others and have less and less visibility of alternate perspectives, the possible trend we are concerned about is towards increasing global fragmentation into large numbers of disjoint groupings, accompanied by erosion of national identities, and weakening of the democratic base. We study these new long-range interactions and their disruptive potential through both historical analysis and modelling of the dynamics, and draw conclusions about the risks and their consequences.

An Historical Overview

Humans are a social species and their societal structures have always been shaped by their communications with each other. Through their interactions they influence each other in many ways – sharing, copying, learning, cooperating, competing and applying pressure – whether forcibly or more subtly – to shape others' behaviours to their own ends. Out of these networks of interactions, human culture evolves and is transmitted, and social structures and processes emerge and are reinforced. Until the emergence of writing such interactions were mostly face to face and immediate. Interactions between people separated in space and/or time were only possible through intermediaries and therefore slow and unreliable, as well as possibly modified by the messengers. It is reasonable to suppose that most of the time the local interactions were the dominant drivers of social processes, with occasional instances of long-range communications providing fresh (and sometimes important) inputs.

The emergence of writing enabled more accurate transmission of messages across long distances and into the future, and increased frequency and impact of long-range one-to-one communications, moderated by the slow rate at which literacy spread for many centuries. One-to-many long-range communications was also feasible if a written message was read out publicly. Thus the influence-reach of a few individuals slowly extended, but it took the invention of the printing press to accelerate the distribution and reach of written information and to stimulate increasing literacy.

Sixteenth century printing innovations introduced a new era of mass communication, permanently altering the structure of society: relatively unrestricted circulation of information and (revolutionary) ideas transcended borders, threatening the power of political and religious authorities, while rising literacy broke the elite's monopoly on education and bolstered the emerging middle class, increasing cultural self-awareness, fostering proto-nationalism and supplanting Latin with local vernacular languages.

In the 19th century industrialised printing by steam-powered rotary presses replaced hand-operated Gutenberg-style presses and the new 'print media' gained rapid ground, being used to both advance the agenda of governments or capitalists, and also enabling pamphleteers to provide some antidote – analogous to today's tweets – to these concentrations of media power.

Twentieth century technologies radically transformed the dissemination of information and culture by introducing both broadcast media – radio and television – and improved one-to-one communication – telephones and telegrams. Nevertheless they still represented a relatively small number of monopolies that were able to control public discourse and the shared narratives, heavily biased towards national and local issues, with only limited channels for individual citizen voices through such means as talkback radio, letters to the editor, and audience and panel discussion television shows. Governments and major political parties were able to address the public through the key news outlets, but minority and fringe views found it more difficult to reach a mainstream audience. Importantly, this meant that a large fraction of the public were exposed to a common pool of information and to a range of views on the issues of the day, including those they did not agree with. But the rise of the internet, personal devices and explosion of computing power changed everything yet again.

The dawn of the internet age brought email and early browsers, and predictions of a socially divisive impact. But in fact costs fell so rapidly that the digital divide that actually emerged was not based on whether people had access but rather on how they used it. There were also fears of political impacts if global information access led to a multiplicity of public spheres and international factions, and a consequent decline in the primacy of the nationstate and in the ability of political parties to act as 'gatekeepers' of national norms and values as they had done till then. And in fact, as users took up new opportunities to connect with others, their sense of identity in their 'real' lives did begin to diverge from the social identities they acquired in these new virtual interaction spaces (*1*), which, unlike the real world, provided both potential anonymity and simultaneous total visibility of others' behaviours, generating powerful new social pressures.

In the last two decades the proportion of the global population with access to the internet has grown to 52%, with early adopters, Europe and North America, at 80% and 88% respectively, and developing regions rapidly catching up (*2*). Internet access also became increasingly ubiquitous with the phenomenal rise of smartphones from 2007 onwards, especially in younger demographics and developing regions. It seems clear that widespread use of mobile devices to access internet services will become the global norm as developing areas catch up and markets saturate.

This period also saw the rise of social media applications, with the daily time spent on social networks worldwide rising from 90 mins in 2012 to 135 mins in 2017, and continuing to rise as mobile internet use increases (*3*).

In summary, globally, more and more people are spending more and more time online, and the fraction of online time that is spent in interactive sites such as social media and online services is also growing.

The Sociobiological Context

Why is this happening so fast and what consequences are there for the structure and functioning of human societies?

To probe these questions more deeply it is necessary to understand how several contributing factors interact, and how in some cases, they accelerate each other's growth in positive feedback cycles.

It is also important to understand some aspects of the context in which these factors exist: a population of actual and potential users who are autonomous agents in their own right, with their own individual desires and purposes, capabilities and limitations. At the same time, from the perspective of service providers, commercial interests and power brokers, they are also markets to be competed for, targets for selling products and services to, and targets for political manipulation.

There are well-documented natural limitations of human cognition and information processing – bounded rationality, limited attention span and working memory, and cognitive biases and heuristics that evolved to reduce the brain's workload in arriving at 'good enough' decisions (*4*). As a consequence, humans are prone to tunnel vision, confirmation bias, susceptibility to believing false information, poor estimation of probabilities and trends, and a host of other fast but faulty thinking patterns, particularly when pressured by ambiguity, too much information and not enough time. In the increasingly information-rich online environment where users are spending more of their time these conditions often hold, but resisting these tendencies requires effortful attention and the ability to tolerate a degree of discomfort in the face of unfamiliar, confusing or challenging information.

In response, and in their ceaseless drive to attract and retain more users, online services and social media were quick to recognise the opportunity to offer users a more enjoyable customised and socialised user experience through personalised content, and through connecting them interactively to other relevant sites and likeminded users. Personalising recommender systems first appeared in the mid-90s, and have rapidly become both more widespread and more powerful ever since, leveraging, and simultaneously driving, the collection of increasing amounts of data about users' browsing histories, which enables them to anticipate with great precision what and who users would want to see more of, and what information they would ignore.

This made it easier for individual users to find material and connections that they were interested in, and therefore accelerated the growth of connected networks of likeminded users and services. It also made it more comfortable for users, reducing both the cognitive workload that sifting through unfiltered news, commentary and other sources would have generated, and the cognitive dissonance of encountering information that might have contradicted their opinions and beliefs.

As these connections grew, virtual communities emerged that were organised around shared interests and views, and significantly, since online interactions are not constrained by geography, much of that growth was fuelled by drawing on long-distance connections from global populations. The result, when viewed on a historical scale, is a sharp rise in the fraction of interactions that are now long-distance rather than local.

This represents a major transformation in the patterns of human interaction, and the speed with which this transformation has occurred is unprecedented and still accelerating. We now explore the consequences of this transformation.

Consequences of Changes in Patterns of Human Interaction

The growing influence of long-range interactions on opinions and behaviours relative to the influence of local interactions presages further major structural changes in how human society is organised globally. The theory of phase transitions in systems consisting of many interacting components suggests that when the influence of long-range interactions starts to dominate that of local interactions, they exhibit phase transitions – systemwide rapid rearrangements of how they are organised, and that, as a result, longer-range structures emerge and persist.

Such phase transitions in human society might manifest in many ways, e.g. large swings in allegiances, social norms, expectations, beliefs and behaviours over relatively short periods of time, and we may well be seeing early signs of just that.

Being connected to larger networks of relevant services and other users with shared interests has in many ways enriched users' online experience, but there is a downside. The very same mechanisms that have benefitted them by creation of a personally tailored information environment, also have the potential to narrow their horizons and isolate them, effectively skewing the information presented to them based on their previous browsing history, often without users' awareness, and creating what has come to be known as filter bubbles (*5*). How filter bubbles affect user behaviour depends on many variable factors, and research evidence at this stage is still contested in the literature, but what is much clearer is how filter bubbles contribute to the emergence of echo chambers, and how echo chambers affect user opinions and behaviours.

The echo chamber effect is defined in the literature as occurring when users encounters *only* beliefs or opinions that coincide with their own, and those beliefs are amplified or reinforced by communication and repetition *inside a closed system* (*6*). These definitions emphasise the isolation of users in an echo chamber from alternate views expressed elsewhere.

But the reality is a little more complex. Opinions and behaviours are influenced through several mechanisms, and their net effect is that users do not need to be completely isolated from alternative perspectives to suffer the effects of being in an echo chamber – the reinforcement and amplification of prior tendencies, coupled with increased hostility to contrary views. One process is group polarisation. When people find themselves in groups of likeminded types, they are likely to move to a more extreme version of their initial positions, through a combination of social influences (the desire to conform and enhance their social reputation in the group they identify with), and the informational deficit of being in a likeminded group (the arguments they are likely to hear in that group are already skewed towards the initial position and hearing more of them tends to push people further in that direction) (*7*). When the information to be assessed is an affective issue to the user, the unconscious processes described as *motivated cognition* (*8*) may kick in to protect their preexisting feelings and beliefs. These include *biased information search* – seeking out (or disproportionally attending to) evidence that is congruent rather than incongruent; *biased assimilation* – crediting and discrediting evidence selectively so as to promote rather than threaten their beliefs; and *identity-protective cognition* – reacting dismissively to information when accepting it would challenge an aspect of their perceived identity. These well-documented processes all conspire to draw users further into patterns of beliefs and attitudes which are resistant to challenge, in spite of including demonstrably false elements.

Of course narrow focus news sources are not a problem if there is a diversity of such sources and users choose a diverse assortment to listen to. But even when sources do produce balanced and diverse offerings, users may still select a subset that they find congruent with their views. On the other hand, if the media's business drivers today are forcing sources that once produced balanced coverage to seek more specialised niches, then it will be more difficult for users to achieve balanced coverage by combining narrow sources.

So while it is true that users can always choose to connect to diverse sources, specifically seek out conflicting views, explore unfamiliar websites, and decline to click on *'recommended for you'* links, the reality is that it may not be so easy to do that, and limited human capacity to deal with the complexity, quantity, and rate of increase, of information they are confronted with, makes following likeminded others, and using the online tools now commonly available in online services and social media apps to filter, personalise, and aggregate information more attractive. In fact, many users have chosen to do just that, thereby further insulating themselves from unwelcome content.

System Dynamics – Impacts on Global Order and Democracy

There are a number of grounds for concern about the impact of these changes.

One is that early hopes for the internet to strengthen democracy are not only dashed but reversed. Instead of the internet being a shared interaction and information space where access is equalised, mutual understanding can flourish and gaps can be bridged, a filtered and customised internet is effectively compartmentalised into a number of domains which have little overlap, and within which users' chances of meeting with unfamiliar ideas are diminished, with fewer rather than more opportunities for constructive engagement and deliberation between different perspectives (*9*).

Warnings about these risks (*10*) were raised as early as 1995, but went largely unheeded, while online services and social media underwent rapid expansion and competed for users. Moreover, the success of the personalisation systems drove more collection and integration of user data, and the utility of these datasets in turn drove development of other ways of exploiting the massive amounts of user data that were now being compiled, privacy concerns notwithstanding. The internet giants such as Microsoft, Google and FaceBook, holding the majority of the user data, found ways to monetise these data assets through targeted advertisements, generating massive revenues for themselves, and accelerating their growth into virtual monopolies. Their success means that their incentives to perpetuate and protect their advantages will now make it very difficult to effectively challenge their *modus operandi*.

The ease with which filter bubbles and echo chambers can be engineered to deliberately manipulate opinions and behaviours means that they will almost certainly be increasingly exploited for marketing, political or even strategic purposes. This is compounded by the ease with which misinformation (when false information is shared, but no harm is meant) and disinformation (when false information is knowingly shared to cause harm) is able to spread, and have impact, online. Retweets and shares can rapidly proliferate information through receptive virtual communities irrespective of its 'truth'. It has been demonstrated that users' opinions are influenced by such content, despite the wide availability of contrary factual information and factchecking services online. This is the *illusory truth cognitive bias* at play, by which users come to believe a message more, the more often it is repeated, even if they initially held a contrary belief. Therefore users are susceptible to being influenced by the most repeated content in echo chambers. If the content is congruent with their prior beliefs then confirmation bias can also augment the effect.

The possibility also exists of lobby groups capturing the support of a larger share of the population through targeting and infiltrating virtual communities organised around a neutral shared interest, winning trust and acceptance by providing desirable content relevant to that interest, and then through the above mechanisms, gradually manipulating the views of its members on the lobby group's core issues.

Another compounding effect is the widely acknowledged general disenchantment with established western governments unable to stem rising inequality, unemployment and national debt, which may increase vulnerability to such political manipulation, increasing ideological polarisation, and enabling demagogues to accrue power, as evidenced by growing populism in western national politics. The interaction with inequality, both actual and perceived, is potentially a very significant modifier of the dynamics. The negative impact of real and perceived inequality on all facets of society is now widely recognised (*11*) and has led to large sections of the community even in wealthy nations feeling increasingly disadvantaged. Echo chambers thrive in such conditions and can in turn drive populist movements to challenge mainstream democratic organisations.

Most significantly, together these factors could create a potentially dangerous positive feedback cycle, running counter to liberal democratic values and norms. Democracy requires an engaged and well-informed electorate to function well. Vested interests can more easily increase their political power by manipulating the electorate to their own advantage, if it is already disengaged and misinformed. Increased concentrations of power in vested interests will probably further increase inequality and thereby further erode the electorate's confidence in the establishment, which feeds back into further disengagement and undermining of the social contract. As a result we also expect that the tendencies already observed towards social fragmentation, alienation from the establishment, distrust of authority and 'out' groups, gullibility to messaging aimed at common affective responses, and vulnerability to manipulation will intensify, particularly in situations where the political environments are conducive.

Unless democratic values and norms can develop to accommodate the emergence of cyberspace, and to effectively counter these risks, alternate possible futures seem grim. We may see a greater concentration of power and greater inequalities that may be temporarily stable with sufficiently 'domesticated' and passive populations, but at the cost of stifled innovation and productivity, and reductions in diversity and abilities to deal with global problems. However the inner tensions of such scenarios imply a precarious stability that could erupt in violent conflict seeded by any number of possible triggers.

The indications are that the underlying drivers of these trends will accelerate for the foreseeable future and that therefore there is a risk of serious instabilities in the international geopolitical order. However it is not so clear how the impacts will play out. For example, mitigating developments may also arise to limit them, and the phase transitions that do occur may lead to quite novel threats, but also to new opportunities.

It is our opinion that in order to remain sustainable, democracy as we know it will need to evolve, and if it is to be effective as well as sustainable, then even more so will it need to change. But it is not at all inevitable that this will happen.

References and Notes

1. M. Lea, R. Spears, Computer-mediated communication, de-individuation and group decision-making. *International Journal Of Man-Machine Studies*, **34**, 283-301 (1991).

2. Global Web Index 2017 p. 38 (https://cdn2.hubspot.net/hubfs/304927/Downloads/Trends-17.pdf), accessed 1 December 2017.
   Internet World Statistics 2017 (http://www.internetworldstats.com/stats.htm), accessed 3 December 2017.

3. M. Meeker, 2017 Internet Trends Report (http://www.kpcb.com/internet-trends), accessed 19 January 2018.

4. D. Kahneman, *Thinking, Fast and Slow*. (Farrar, Straus and Giroux, New York, 2011)

5. D. Spohr, Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review* **34**, 150-160 (2017).

6. https://en.oxforddictionaries.com/definition/echo_chamber
   https://en.wikipedia.org/wiki/Echo_chamber_(media)

7. M. S. Levendusky, Why Do Partisan Media Polarize Viewers?. *American Journal of Political Science*, **57**, 611-623 (2013).

8. D. M. Kahan, Misconceptions, Misinformation, and the Logic of Identity-protective Cognition. *The Cultural Cognition Projext, Working Paper No 164*. (Yale University, New Haven, 2013).

9. http://insidestory.org.au/in-a-bubble-on-the-web/

10. M. Van Alstyne, E. Brynjolfsson, Could the Internet Balkanize Science?. *Science*, **274**, 1479-1480, (1996).

11. https://www.imf.org/external/pubs/ft/sdn/2015/sdn1513.pdf

Complex software projects and the problem of estimating the time to create them

Diplom-Informatiker Maximilian Hirner*

*E-mail: fractals.procgen@gmail.com

Since the time when software projects became non-trivial, the estimating of the time they will take from start to finish has been a major, if not the biggest problem of the software industry. Many different approaches have been tried, but don't seem to have lead to a solution. The author's hope: By considering the unsolved problem as a clouded kind of fractal – in the sense that problems tend to be "hairy", containing unforeseen sub-problems and sub-sub-problems – we can understand the deeper reason for software projects taking too long.

Note: In most complex software projects, various problems with intra-team communication are the main reason for schedule slips. To exclude the influence of these specific reasons from the analysis, only one-programmer software projects were considered.

Introduction

Since at least the birth of the author, IT professionals have pondered the problem of software projects taking much longer than estimated. (*1*) Since differently from e.g. construction business most of the costs of software projects are caused by the wages for the programmers, a reduction of the number of man-months would help greatly to save costs – if this is possible.

Programmers are said to be optimists. (*1*)[p.14] Possibly they have to be, because otherwise many projects (even those the programmers are truly dedicated to) wouldn't even have been started.

The basic idea this text will cover: Estimating the time a certain software project will take is hard because complex projects are "fractal-like", i.e. if depicted in an abstract way as a tree of tasks, sub-tasks, sub-sub-tasks etc, the resulting shape will be a kind of fractal. – While not every complex problem would be this kind of fractal, every fractal-like problem would have to have a certain level of complexity.

While this idea will not help us to get better estimations for the duration of software projects (at least not immediately), at the very least it will give us a better explanation for their missed schedules.

Old approaches

There are several different approaches to the question why the problem of software projects taking longer than expected continues to happen:

1. Too many programmers working on a project lead to a nonlinear (i.e. greater than linear) growth of communication channels, hence Brooks's law: ”Adding manpower to a late software project makes it later.” (*1*)[p.25]

2. The complexity of projects simply becomes too big to handle. (*2*)

3. Software projects, like many other man-made things, will have a ”life expectation” that grows for more than one day for each day people are working on them. Meaning: If a project already took 79 days and isn't finished yet, it will probably take 25 more days. If it isn't finished after 100 days, expect it'll take 89 more days. And if it isn't finished after 600, we'd even have to expect it'll take 1590 more days. (Numbers taken from (*3*)[p.200].)

Approaches 2 and 3 explain the observed facts in a very rough way, but the deeper reasons for the failings of projects stay opaque. Approach 1 gives a concrete explanation, although it begs the question whether this is the only reason for projects taking too much time.

Analyzed projects

As said in the abstract, only one-man projects will be considered for this paper – to be precise, projects done by the author. This condition was introduced to exclude the influence of intra-team communication. This restricts the research to smaller projects (10,000 to 50,000 LoC, including scripts that were written as minor tools for the purpose of these projects), but may still help to answer the basic question of the paper. Also, it avoids legal problems with the author's former employers.

Whenever these projects will be referred to in the paper, the projects in question will be code-named (for reason of simplicity) Project 1, 2 and 3 respectively.

The analysis of a software problem

For this, we can take a similar way as in complexity science: ”Scientists look at something and want to understand what it does, and how it does it. One of the key observations about the world is that everything is made up of parts. So reasonably enough, scientists, trying to figure out what the object does, work to figure out what the parts do. When we look at one of the parts, we realize that it too is made up of parts. The next step, therefore, is to look at the parts that make up the part to figure out what they do. This continual breaking down of parts into their component parts progresses until we forget what it was that we were trying to do in the first place!" (*2*)[p.23]

Initially, we may consider a software project as one big task. Next, we would develop a mental model for it. Software to be installed on PCs may follow e.g. an MVC architecture. This big task has to be broken into sub-tasks, as the first step. Web-based software which is based on a LAMP server e.g. might be divided into these immediate sub-tasks:

- Configure the Linux OS of the web server
- Configure the web server software (including security)
- Create and configure the database
- Write several scripts (core of the software project)
- Create the GUI

Afterwards, the individual sub-tasks will have to be broken into sub-sub-tasks, and so on.

Encountered problems in practice

One important problem encountered in Project 3 was the fact that part of the original job definition was vague, along the line of "take this text corpus and remove markup mistakes and other common typos, to prepare it for further processing by the computer". While this is only a part of the whole project, it demonstrates the complexity well enough, so no further problems will be considered in this text.

Said text corpus consisted of several thousand text files in a common markup language (sourced from the internet) created by many different human authors of varying qualification. Hence one had to expect that the text (especially considering the markup) would be flawed and contain many errors.

Some sub-problems involved with the handling of the text corpus:

- Broken HTML entities (e.g. "&" rendered as "&", "&" or "&amp")
- Mojibake, e.g. garbled non-ASCII text because of different encodings (UTF-8, ISO 8859-1) – "£" might become "Â£"
- Broken Markup ("[[Text]]" or "{{Text}}" rendered with too many or too few brackets/braces)

- Wrong kind of markup (HTML tags or BBCode in wiki markup or vice versa)
- Markup nested the wrong way (e.g. ”<i>Some text here</i>” instead of ”<i>Some text here</i>”)

The project on an abstract level

When you plan a software project, you create a mental model of the finished software. But any model, even a good model free of flaws, is by necessity a simplification of reality.

In bigger projects e.g. involving big text corpora (as was the case with Project 3), it's impossible to achieve an exhaustive solution, since this would make it necessary to read and error-check thousands of text files. Hence, errors of the same kind were only corrected if at least one of this kind was discovered in a random sample. Since the absence of proof for further errors doesn't equal proof of absence, the existence of further errors was possible (and during later work, often enough proven).

Let's say we had a text corpus T (again, a number of text files), and { A, B, C, X, Y, Z } were given non-empty strings, with X, Y and Z being all distinct from each other. We wanted to replace all occurrences of string A in T by string X. Only later however we find out that in cases where string A is followed by string B, we have to replace the concatenation of A and B by string Y instead. And even later we may find out that there's an exception to the exception, having to replace ABC by Z. This abstraction e.g. fits what was a part of the work on Project 3.

Development of the mental model of the project

(The idea covered by this chapter was inspired by the mental model of happiness in (*4*))

Explanation: In the graphics shown in this paragraph, the color black symbolizes ”work that will have to be done”, the color white ”work that won't have to be done”, and the color grey ”work that might have to be done” – or rather, that what is uncertain and not covered by the mental model yet.

A project might start with a vague idea, symbolized by this:

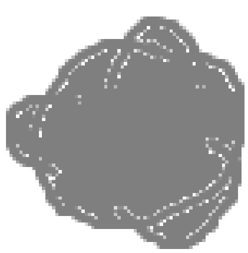


Figure S1

As soon as there is a rough sketch of the work to do, we will get this:

Figure S2

As soon as there is a sketch of the immediate sub-tasks to do, we will get this:

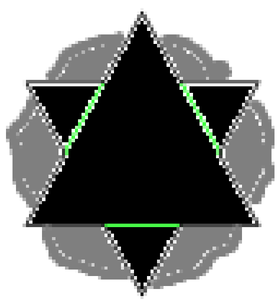


Figure S3

This recursive process may go on for as many levels as you like.

(Interpretation: Each triangle is standing for a simple, crisp task that can be solved in a straightforward way. Tasks like ”Open a specified file and process the content with function X” or ”If the user clicks button Y, call function Z”. The bread and butter of every programmer. The assumed scalability of these tasks makes sure that a task will work out in principle, no matter whether it deals with one kB or one GB of data.)

It becomes obvious not so much that the number of sub-tasks is growing exponentially (depending on the complexity level of project, this might not be the case), but that because of scalability the small tasks will create a similar amount of work as the big ones.

Trying to measure complexity

If the theoretical solution to our problem can be abstracted as a tree, as long as we are in the planning stage we won't know the whole tree. By necessity we have to deal with subsets.

We have seen: A project can be depicted as a tree where each branch is standing for a sub-task, sub-sub-task etc. Now it is obvious: If the planners forget about a certain branch, they will also overlook all the sub-branches etc. of said branch. Mistakes like these can happen on any possible complexity level – without further research, it's too early to tell whether there are variances on different levels.

Let us assume that on each level, each branch will split up into three sub-branches, but the planners will consider only two of them, for lack of imagination.

Then we can calculate for up to five complexity levels:

L	C	R	P	O
0	1	1	0	0
1	2	3	33.3	50
2	4	9	55.6	125
3	8	27	70.4	212.5
4	16	81	80.2	406.3
5	32	243	86.8	659.4

Legend:

L	Complexity level
C	Considered branches at the lowest level
R	Real number of branches at the lowest level
P	Percentage of branches not planned for (%)
O	Overhead assuming each branch will take the same time (%)

(*1*)[p.88] tells us indeed that the exponent (time needed relatively to the size of the program) is about 1.5!

Conclusion

While a few hobby projects don't tell us too much about the software industry in general, this short paper hopefully demonstrates how the natural complexity of some projects enlarges planning errors – the higher the complexity, the more. And while the measure of complexity from the chapter "Development of the mental model of the project" wasn't based on measuring many real world projects, the author would like to point out that it was conceived before the lecture of (*1*).

References

1. "The Mythical Man-Month", Frederick P. Brooks jr., Addison-Wesley, 1975, ISBN 0-201-00650-2

2. ”Making Things Work”, Yaneer Bar-Yam, NECSI Knowledge Press, Boston, 2004, ISBN 0965632822

3. ”Der Schwarze Schwan” (German translation of ”The Black Swan”), Nassim Nicholas Taleb, Hanser, Munich, 2008, ISBN 3446415688

4. ”The Society of Mind”, Marvin Minsky, Simon & Schuster, New York, 1986, ISBN 0-671-60740-5

Supplementary materials

Figures S1 to S3

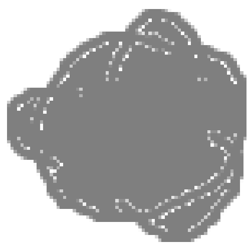


Figure S1 A project might start with a vague idea, symbolized by a grey cloud

Figure S2 As soon as there is a rough sketch of the work to do, we will get this

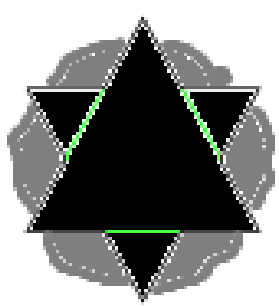


Figure S3 As soon as there is a sketch of the immediate sub-tasks to do, we will get this

Critical dynamics in the Belousov−Zhabotinsky reaction

Harold M Hastings[1,2], Richard J Field[3], Sabrina G Sobel[2] and Maisha Zahed[1]

[1] Bard College at Simons Rock, Great Barrington, MA 01230, USA

[2] Hofstra University, Hempstead, NY 11549, USA

[3] University of Montana, Missoula, MT 59812 USA
hhastings@simons-rock.edu

Abstract. The dynamics of the Belousov−Zhabotinsky (BZ) reaction, the prototype oscillatory chemical reaction, has also been studied as an analog of the dynamics of neurons and the cardiac electrical system [1,2]. Recent analogies with neurons include implementation of gates [3] and excitatory and inhibitory connections [4]. Here we report on an experimental study of the onset of oscillations in the BZ reaction with a variety of catalysts and describe analogies with the onset of seizures in the brain, characterized as a "dynamical disease", c.f. [5], and references therein).

Keywords: Excitable medium, Hopf bifurcation, FitzHugh-Nagumo, dynamical disease

1. Introduction

The Belousov−Zhabotinsky (BZ) reaction is the prototype chemical excitable medium [6-17]. An excitable medium is a system in which sufficiently large perturbations generate activations before returning to steady state. Excitable media support traveling waves of activation and can also display auto-oscillatory behavior. Biological examples include the cardiac electrical system and neural nets media; there are many analogies between BZ dynamics and the dynamics of biological excitable media [1,2].

Unstirred, ferroin/ferriin ($Fe(phen)_3^{2+/3+}$)-catalyzed BZ reaction media in a thin (~ 2D) layer in a Petri dish typically evolve to a near-critical red, reduced, low [ferriin] state during an induction period of several minutes [7,10]. After this induction period, one sees the apparently "spontaneous" formation of target patterns of outward moving waves of oxidation generated by "pacemakers" oscillating between a blue, oxidized, high [ferriin] state and a red/reduced state, which appear to be generated by heterogeneities, fluctuations, or a combination [15-20]. In contrast, typical manganese ($Mn^{2+/3+}$) catalyzed reaction media, c.f. [10], display phase waves or bulk oscillations

The goal of this research is to experimentally explore these differences, better understand the onset of oscillations, and explore analogues with epilepsy, a dynamical disease of the brain [5,21,22]. A variety of abnormal physiological states, including cardiac arrythmias as well as epilepsy, have been characterized as dynamical disease, c.f. [5] and references therein, whose control and treatment requires control of global dynamics.

2. Ordinary differential equation models for BZ reaction and neuronal dynamics

Field, Körős and Noyes [8] developed a skeleton reaction mechanism (FKN) for the BZ reaction, which led to the Oregonator model, a system of three differential equations that captures essentially all the qualitative behavior of the BZ reaction [9].

.

$$
\begin{aligned}
dx/dt &= k_3[BrO_3^-][H^+]^2 y - k_2[H^+]xy + k_5[BrO_3^-][H^+]x - 2k_4x^2 \\
dy/dt &= -k_3\left[BrO_3^-\right]\left[H^+\right]^2 y - k_2\left[H^+\right]xy + 1/2 f k_c[MA]z \qquad (1) \\
dz/dt &= 2k_5[BrO_3^-][H^+]x - k_c[MA]z
\end{aligned}
$$

Here x, y, and z, respectively, correspond to $[HBrO_2]$, $[Br^-]$ and $2[M^{(n+1)+}]$, where $M^{(n+1)+}$ denotes the oxidized form of the metal-ion catalyst, e.g., ferroin or Mn^{3+}, and MA represents all forms of malonic acid, including bromomalonic acid and di-bromomalonic acid. See also [9,10,12,14].

The effects of the concentrations $[H^+]$, $[BrO_3^-]$ and $[MA]$ and the stoichiometric factor *f* on the dynamics are most easily by rescaling the dynamic variables in Oregonator model as in Tyson [10] and then adiabatically eliminating the very fast variable corresponding to the rescaled $\left[Br^-\right]$. We obtain the following form of the Field-Noyes equations, c.f. [23].

$$
\begin{aligned}
dx/dt &= \left(k_3[H^+][BrO_3^-]\right)\Big(x(1-x) - fz\left(x-q\right)/\left(x+q\right)\Big) \qquad (2) \\
dz/dt &= \left(k_c[MA]\right)(x-z)
\end{aligned}
$$

Equations (2) are a relaxation oscillator, consisting of a fast activator (x = rescaled $[HBrO_2]$) and a slow inhibitor (z = rescaled $[M^{(n+1)+}]$), whose dynamics depends upon the effects of the inhibitor on activator dynamics (given by the stoichiometric factor *f* which corresponds to amount of Br^- produced by reduction of $M^{(n+1)+}$. The Showalter-Noyes [24] product $[H^+][BrO_3^-]$ determines the fast (activator x) time scale, and $[MA]$ determines the slow (inhibitor z) time scale, and f a (stable or unstable) steady state of eqs. (2), independently of $[H^+][BrO_3^-]$ and $[MA]$ in rescaled coordinates, c.f. [10,23,25].

Boissonade and de Kepper proposed a generic two-variable (BdK) model [26,27] for chemical excitable media (we write z in place of y, following eqs. (2))

$$
\begin{aligned}
\frac{dx}{dt} &= -\left(x^3 + \mu_0 x + \lambda\right) - kz \qquad (3) \\
\frac{dz}{dt} &= (1/\tau)(x-z)
\end{aligned}
$$

In both eqs. (2) and (3) the *z*-nullcline is the same, the *x*-nullcline has one local minimum and one local maximum, although it has a vertical asymptote only in eqs. (3).

Finally, the FitzHugh-Nagumo neural model [28,29] is represented by eqs. (4):

$$\frac{dv}{dt} = \phi(v) - w + I$$
$$\phi(v) = v(v - a)(1 - v) \tag{4}$$
$$\frac{dw}{dt} = \varepsilon\left(v - \gamma w - \beta\right)$$

The membrane potential v in the FHN model corresponds to the activator ($HBrO_2$) concentration x in the BdK analogue of the Oregonator; similarly the gate variable w in the FHN model corresponds to the inhibitor (oxidized catalyst) concentration z in the BdK analogue of the Oregonator. The FHN model is readily transformed into a corresponding BdK model through a sequence of linear transformations [23]. The dynamics of the FHN model depends upon where the nullclines cross and the ratio of fast to slow time scales $\tau = 1/\varepsilon$ corresponding to the ratio $[H^+][BrO_3^-]/[MA]$ in the dynamics of the Oregonator.

We next explore the dynamics of the spatially extended BZ reaction systems experimentally.

3. Experimental

Methods. Recipes were modified from [11]. All reagents were used as received. We used a total volume of 13 mL in a 90 mm Petri dish, yielding a depth of 2.0mm. First, Reagent mixture A 0.25M $KBrO_3$: 7.5mL, 0.50M KBr: 1.0 mL, 3.0M H_2SO_4: 2.5mL, 0.50M MA, 1.0mL, for a total volume of 12mL was stirred and allowed to equilibrate (FKN reactions [8] FKN1-FKN5 and FKN8), thus essentially completing the bromination of some of the MA to bromomalonic and dibromomalonic acids, c.f. [30]. A catalyst, one or a mixture of 0.025M ferroin, 0.50M $MnSO_4$, 0.50M $CeSO_4$, for a total volume of 1.0mL was added and the mixture was again stirred. The difference in catalyst concentrations is due to differences in their redox potentials. Although data on redox potential varies, the following patterns hold:

(1) The ferroin/ferrin redox couple is significantly weaker than the Mn^{+2}/Mn^{+3} couple (1.06V vs. 1.51 V in 1M $HClO_4$ [31] and the Ce^{+3}/Ce^{+4} couples (1.61 V with no added acid, compared with 1.14V for ferroin/ferriin [31]).

(2) The redox potential of BrO_2 + H^+ /$HBrO_2$ couple is comparable to that of Mn^{+2}/Mn^{+3} and Ce^{+3}/Ce^{+4} (1.49 V, 20°C, in 1M H_2SO_4 [32]).

There are significant effects upon the catalytic oxidation of all forms of MA (Process C, reactions FKN 9 and 10, c.f. Scott (1985)) and the reduction of $M^{+(n+1)}$ (recycling the oxidized catalyst) in Process B of FKN dynamics, reactions FKN 5 and 6:

[FKN 5] $BrO_3^- + HBrO_2 + H^+ \rightarrow 2\ BrO_2 + H_2O$ (rate limiting)
[FKN 6] $BrO_2 + M^{+n} + H^+ \rightarrow HBrO_2 + M^{+(n+1)}$

(1) FKN6 is highly energetically favored for ferriin/ferroin, close to neutral for other catalysts. Thus a higher concentration of Mn or Ce is required to drive FKN6 forward.
(2) Ferriin cannot oxidize MA itself, only its brominated forms [33]. Thus relatively more Br^- is produced in Process C, reduction of $M^{+(n+1)}$, than with other more electropositive catalysts, as described mathematically by an increase in f in eqs. (1) and (2).
(3) Ferroin is readily oxidized by Ce and Mn at rates much faster than the oxidation of MA [31,34]. Since ferriin, the result of this oxidation, will only oxidize brominated malonic acid, a ferroin perturbation also increases f.

Mixed catalyst results. Figure 1 shows the results of experimental interpolation between the catalysts ferroin/ferriin and $Mn^{2+/3+}$. The ferroin-catalyzed reaction shows the usual target waves, and the Mn-catalyzed reaction shows phase waves. At 20-80% ferroin/80-20% Mn, one sees target patterns in an oscillating (phase wave) background. (Small amounts of ferroin (e.g. 2%) do not affect the dynamics of Mn-catalyzed media, c.f. [35].)

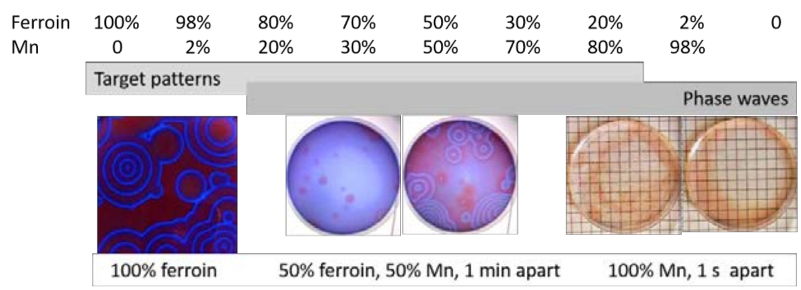


Fig. 1. Interpolating between ferroin/ferriin-catalyzed and $Mn^{2+/3+}$-catalyzed reactions. Percentages refer to fractions of 0.025M ferroin and 0.50M $MnSO_4$ solutions in 1mL of total catalyst solution. For example at ferroin 40%/Mn 60%, 1 mL a catalyst solution of 0.010M ferroin and 0.30M $MnSO_4$ is used in a total reaction volume of 13mL.

Perturbations. We also perturbed BZ reaction media by adding small amounts (0.3μL) of dilute $AgNO_3$ solutions to explore how near the media were to criticality (and thus oscillations). The smallest perturbation causing a change was 10^{-8}M added to the ferroin-catalyzed reaction, slightly reducing [Br^-], generating a local excitation and outgoing waves of oxidation, c.f. [30,36]. As a control, smaller concentrations did not generate a local excitation. Similar additions of ferroin to the Mn-catalyzed reaction and Mn to the ferroin-catalyzed reaction generated phase waves and a local excitation and outgoing waves of oxidation respectively.

Additional results, interpretation. Additional experimental results will appear elsewhere [37]. The above results can be understood in terms of the following bifurcation diagram, Fig. 2, based upon [25].

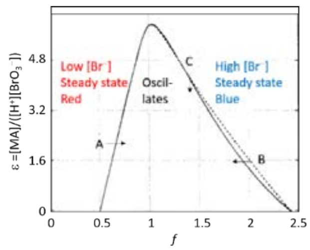


Fig. 2. Bifurcation diagram of the Oregonator, reprinted with permission from Mazzotti, M., Morbidelli, M., Serravalle, G. J. Phys. Chem., 99, 4501-4511 (1995), Figure 1 [25]. Copyright 1995 American Chemical Society. The y-axis is rescaled and relabeled following [23]. Arrows indicate bifurcations: A. Adding ferroin to the Mn-catalyzed reaction increases the stoichiometric factor f since some Mn^{3+} oxidizes ferroin to ferriin, which cannot oxidize MA. B. Adding Mn to the ferroin-catalyzed reaction likely adds a small amount of the impurity Mn^{3+} which oxidizes MA, reducing f. C. Increasing the ratio of the fast to slow time scales can also trigger oscillations, see below.

4. FHN dynamics, dynamical disease, epilepsy

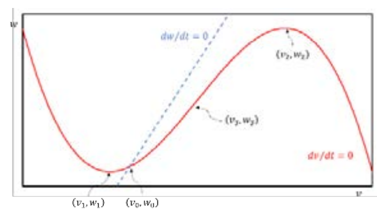


Fig. 3. Phase portrait of FitzHugh-Nagumo dynamics.

Fig. 3, above is a phase portrait of FHN dynamics. There is one equilibrium $(v_0,\ w_0)$, which is unstable if the nullclines intersect at a point between the local minimum $(v_1,\ w_1)$ and local maximum $(v_2,\ w_2)$ of the v-nullcline, and $\tau = 1/\varepsilon$ is sufficiently large (depending upon $\phi'(v)$); otherwise the equilibrium is stable. Moreover, there is one point of inflection at $v = v_3$ with $v_1 < v_3 < v_2$, below which $\phi'(v)$ increases with increasing v. For $v_1 < v < v_3$ there is a subcritical Hopf bifurcation as the bifurcation parameter

$$\phi'(v) - \varepsilon\gamma \tag{3}$$

increases through 0.

Since $f'(v)$ is an increasing function of v between the local minimum and point of inflection (v_3, w_3), in this range, the Hopf bifurcation can be triggered by increasing the resting (steady state) membrane potential (v_0), or speeding up the fast time scale $\tau = 1/\varepsilon$ (increasing excitability), or a combination.

Connections with epilepsy. These factors correspond to two potential triggers of epileptic seizures: disturbances in intracellular or extracellular ion concentration increasing the resting potential and increases in excitability caused by overly high [glutamate] or overly low [GABA], c.f. [38-40] and references therein.

Analogies with BZ dynamics. Increasing the resting potential in FHN dynamics corresponds with increasing steady state [$HBrO_2$] in the BZ reaction, which corresponds to decreasing steady state [Br^-] through decreasing the stoichiometric factor *f* or decreasing exogeneous Br^- production. Similarly, increased excitability corresponds to a faster "fast activator time scale," from an increased Showalter-Noyes product [H^+] [BrO_3^-] in the BZ reaction, and decreased inhibition to a slower "slow inhibitor time scale," from decreased [MA] in the BZ reaction c.f. [23]. Thus loosely, comparing neuronal dynamics with BZ dynamics, [H^+][BrO_3^-] ~ [glutamate] and [MA] ~ [GABA].

It is thus possible that explorations of the dynamics of spatially extended (unstirred) BZ systems, especially the onset of oscillations and the excitable/oscillatory transition through a Hopf bifurcation [15,16,41] may lead to interesting questions about bifurcations underlying the onset of epilepsy [42-44].

Limitations. First, this analogy does not capture significant differences between BZ and brain dynamics, and in fact, does not describe the full spatio-temporal dynamics of either system.

5. Discussion

We have briefly explored the effects of catalyst mixture and small perturbations on critical dynamics in the BZ reaction, and used equivalence of the generic BdK model for chemical excitable media and the FitzHugh-Nagumo neuronal model to draw analogies with the transition from normal brain dynamics to epilepsy.

References

1. Glass, L., Mackey, M.C.: From Clocks to Chaos: The Rhythms of Life. Princeton Univ. Press, Princeton, NJ, USA (1988).
2. Winfree, A.T.: The Geometry of Biological Time, 2nd ed., Springer, New York (2001).
3. Adamatzky, A.: Fredkin and Toffoli gates implemented in Oregonator model of Belousov–Zhabotinsky medium. Int. J. Bif. Chaos, 27, 1750041 (2017).

4. Litschel, T., Norton, M.M., Tserunyan, V., Fraden, S.: Engineering reaction–diffusion networks with properties of neural tissue. Lab on a Chip 18(5), 714-722 (2018).
5. Glass, L.: Dynamical disease: Challenges for nonlinear dynamics and medicine. Chaos 25(9), 097603 (2015).
6. Belousov, B.P.: In: Field, R.J., Burger, M. (eds.), Oscillations and Traveling Waves in Chemical Systems, pp 605-613. Wiley-Interscience: New York, 1985;
7. Zaikin, A.N.; Zhabotinsky, A.M.: Concentration wave propagation in two-dimensional liquid-phase self-oscillating system. Nature, 225(5232), p.535 (1970).
8. Field R. J., Körös E, Noyes R.M. Oscillations in chemical systems. II. Thorough Analysis of Temporal Oscillation in the Bromate-Cerium-Malonic Acid System. J. Am. Chem. Soc. 94, 8649-8664 (1972).
9. Field, R. J., Noyes, R. M.: Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction. J. Chem. Phys. 60, 1877-1884 (1974).
10. Tyson, J.J.: Scaling and reducing the Field-Koros-Noyes mechanism of the Belousov-Zhabotinskii reaction. J. Phys. Chem. 86(15), 3006-3012, (1982).
11. Shakhashiri, B.Z.: Chemical Demonstrations, vol 2, Univ. Wisconsin Press, Madison, Wisc., USA, pp 273 – 299 (1985).
12. Field R.J., Försterling H-D.: On the oxybromine chemistry rate constants with cerium ions in the Field-Körös-Noyes mechanism of the Belousov-Zhabotinskii reaction: The equilibrium $HBrO_2 + BrO_3^- + H^+ \Leftrightarrow 2BrO_2^{\bullet} + H_2O$. J. Phys. Chem. 90, 5400-5407 (1986).
13. Scott, S.K.: Oscillations, Waves and chaos in chemical kinetics. Oxford University Press, New York (1994).
14. Hegedus, L., Wittman, M., Noszticzius, Z., Yan, S., Sirimungkala, A., Försterling, H-D., Field, R.J.: HPLC analysis of complete BZ systems. Evolution of the chemical composition in cerium and ferroin catalysed batch oscillators: Experiments and Model Calculations. Faraday Discussions 120, 21-38 (2001).
15. Hastings H. M.; Field R. J.; Sobel, S. G. Microscopic fluctuations and pattern formation in a supercritical oscillatory chemical system. J. Chem. Phys. 119, 3291-3296 (2003).
16. Sobel, S.G., Hastings, H.M., Field, R.J.: Oxidation state of BZ reaction mixtures. J. Phys. Chem. A, 110(1), 5-7 (2006).
17. Bugrim, A.E., Dolnik, M., Zhabotinsky, A.M., Epstein, I.R.. 1996. Heterogeneous sources of target patterns in reaction-diffusion systems. J. Phys. Chem. 100(49), 19017-19022.
18. Dutt, A.K., Menzinger, M.: Stirring and mixing effects on chemical instabilities: bistability of the bromate/bromide/cerium (3+) system. J. Phys. Chem. 94(12), 4867-4870 (1990).
19. Vidal C., Pagola, A.: Observed properties of trigger waves close to the center of the target patterns in an oscillating Belousov-Zhabotinskii reagent. J. Phys. Chem. 1989, 93(7), 2711-2716 (1989).
20. Zhang, Y.-Z., Förster, P., Ross, J.: Origin of spontaneous wave generation in an oscillatory chemical system. J. Phys. Chem. 96, 8898-8904 (1992).
21. Bélair, J., Glass, L., an der Heiden, U., Milton, J.: Dynamical disease: identification, temporal aspects and treatment strategies of human illness. Chaos 5(1), 1-7 (1995).
22. Milton, J.G.: Epilepsy: multistability in a dynamic disease, pp. 374-386. Cambridge Univ. Press, Cambridge, UK (2000).
23. Hastings, H.M., Field, R.J., Sobel, S.G., Guralnick, D.: Oregonator scaling motivated by the Showalter–Noyes limit. J. Phys. Chem. A 120(41), 8006-8010 (2016).
24. Showalter, K., Noyes, R.M.: Oscillations in chemical systems. 15. Deliberate generation of trigger waves of chemical reactivity. J. Am. Chem. Soc. 98, 3730-3731 (1976).

25.Mazzotti, M., Morbidelli, M., Serravalle, G.: Bifurcation analysis of the Oregonator model in the 3-D space bromate/malonic acid/stoichiometric coefficient. J. Phys. Chem. 99(13), 4501-4511 (1995).
26.Boissonade, J., De Kepper, P.: Transitions from bistability to limit cycle oscillations. Theoretical analysis and experimental evidence in an open chemical system. J. Phys. Chem. 84(5), 501-506 (1980).
27.Epstein, I.R., Showalter, K.: 1996. Nonlinear chemical dynamics: oscillations, patterns, and chaos. J. Phys. Chem. 100(31), 13132-13147 (1996).
28.FitzHugh R.: Mathematical models of threshold phenomena in the nerve membrane. Bull. Math. Biophysics. 17, 257-278 (1955).
29.Nagumo J., Arimoto S., Yoshizawa S.: An active pulse transmission line simulating nerve axon. Proc IRE. 50, 2061–2070 (1962).
30.Hastings, H.M., Sobel, S.G., Field, R.J., Bongiovi, D., Burke, B., Richford, D., Finzel, K. and Garuthara, M.: Bromide control, bifurcation and activation in the Belousov−Zhabotinsky Reaction. J. Phys. Chem. A, 112(21), 4715-4718 (2008).
31.Diebler, H., Sutin, N.: The kinetics of some oxidation-reduction reactions involving manganese (III). J. Phys. Chem. 68, 174-180 (1964).
32.Armstrong, D.A., Huie, R.E., Koppenol, W.H., Lymar, S.V., Merényi, G., Neta, P., Ruscic, B., Stanbury, D.M., Steenken, S., Wardman, P.: Standard electrode potentials involving radicals in aqueous solution: inorganic radicals (IUPAC Technical Report). Pure Appl. Chem. 87, 1139-1150 (2015). https://www.bnl.gov/isd/documents/92710.pdf, p. 122, last accessed 2018/4/26.
33.Rovinskii, A.B., Zhabotinskii, A.M.: Mechanism and mathematical model of the oscillating bromate-ferroin-bromomalonic acid reaction. J. Phys. Chem. 88(25), 6081-6084 (1984).
34.Dulz, G.,Sutin, N.: 1963. The kinetics of the oxidation of iron (II) and its substituted tris (1, 10-phenanthroline) complexes by cerium (IV). Inor. Chem. 2, 917-921 (1963).
35.Manz, N., Steinbock, O.: 2006. Propagation failures, breathing pulses, and backfiring in an excitable reaction-diffusion system. Chaos 16(3), 037112 (2006).
36.Noyes, R.M., Field, R.J., Försterling, H.D., Koros, E. and Ruoff, P.: Controversial interpretations of silver (1+) perturbation of the Belousov-Zhabotinskii reaction. J. Phys. Chem. 93(1), 270-274 (1989).
37.Zahed, M. and Hastings, H.M. Experimental study of critical dynamics in the BZ reaction. In preparation.
38.Bromfield, E.B., Cavazos, J.E. and Sirven, J.I., 2006. An introduction to epilepsy. American Epilepsy Society, West Hartford, CT, USA (2006). https://www.ncbi.nlm.nih.gov/books/NBK2510/, last accessed 2016/5/3.
39.Raimondo, J.V., Burman, R.J., Katz, A.A., Akerman, C.J.: 2015. Ion dynamics during seizures. Frontiers Cellular Neuroscience, 9, 419 (2015).
40.Di Garbo, A., Barbi, M., Chillemi, S.: The role of glutamatercic and GABAergic synapses on the dynamics of neural networks: How they impact the transition to seizure?. Neurocomputing. 114, 15-23 (2013).
41.Baer, S.M., Gaekel, E.M.: Slow acceleration and deacceleration through a Hopf bifurcation: Power ramps, target nucleation, and elliptic bursting. Phys. Rev. E, 78(3), 036205 (2008).
42.Jirsa, V.K., Stacey, W.C., Quilichini, P.P., Ivanov, A.I. and Bernard, C.: On the nature of seizure dynamics. Brain 137(8), 2210-2230 (2014).

43. Wang, S.J., Ouyang, G., Guang, J., Zhang, M., Wong, K.M., Zhou, C.: Stochastic oscillation in self-organized critical states of small systems: Sensitive resting state in neural systems. Phys. Rev. Lett. 116(1), 018101 (2016).
44. Liu, S., Wang, Q.: Transition dynamics of generalized multiple epileptic seizures associated with thalamic reticular nucleus excitability: A computational study. Comm. Nonlinear Science Numerical Simulation, 52, 203-213 (2017).

Dynamics of Lifespan Evolution

Ryan W. Gryder,[1*] Mark A. Kon,[1,2]

[1]Department of Mathematics and Statistics, Boston University,
Boston, MA 02215, USA
[2]Department of Mathematics, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA

*To whom correspondence should be addressed; E-mail: rwgryder@bu.edu.

The evolutionary purpose of aging and disease has been and continues to be contested among scientists today. Here we consider carefully the view that a species-specific optimal lifespan is programmed into the genome and is enforced by disease. Our essential argument is that evolution operates at the individual level on a short time scale as well as at the group level on a longer time scale, and the result is an optimal lifespan that balances the good of the individual against the good of the species. We present a dynamical partial differential equation model in genome space and a computer simulation of the time evolution of a population genome under the influence of a fitness function. The results demonstrate an advantage of a shorter lifespan in maintaining population size and robustness in the long run: with a shorter lifespan, a population is able to evolve quickly enough in response to its changing environment. We show that mechanisms enforcing a limited lifespan can be encoded with meta-evolutionary mechanisms, i.e., processes involving the evolution of evolution. Such mechanisms reduce mathematically to 'dynamics of dynamics': meta-dynamics of evolution represented by dynamical systems whose parameters change slowly (adiabatically) in time. We also consider the role of disease manifolds in genome space in enforcing optimal lifespans and consider diseases as adaptive mechanisms with directed evolutions. Specifically, we discuss the evolutions of cancers as potential adaptive mechanisms related to lifespan control.

Introduction

Scientific research aimed at understanding the biological mechanisms behind aging and disease continues to grow. Much of this is dedicated to locating specific aging or disease genes within our genome and those of closely related species. The question posed in these studies is often, 'how can we prevent this disease?' or 'how can we slow down the aging process and live longer?' However, when one considers disease and aging from an evolutionary biology perspective, a more fundamental question regarding these biological processes arises: why should aging and disease exist at all if they are in direct opposition of the biological goal of passing on genes in maximal multiples? Some answers to this question include the argument that after prime reproductive age, an organism has not much to offer to the well-being of the population and withers and dies as a result. Then comes an equivalent question: why doesn't evolution lead to a longer reproductive period? Many answers related to aging are not satisfactory from an evolutionary point of view. An alternative claim maintains that a species-specific lifespan is genetically programmed for the long-term benefit of the species (*1*). We extend this claim by modeling its genomic instantiation mathematically, and we consider related enforcement mechanisms in the form of diseases.

Individual vs. group. A key component of the debate on whether finite lifespan is programmed is related to the question of whether group selection (toward finite lifespan) can dominate individual selection (toward a longer one). Game theorists continue to study the balance between individual behaviors that are cooperative (benefit the group) versus competitive (benefit themselves) (*2*). It is within this framework that we will consider the evolutionary role of aging and disease. At the individual level, it is clear that aging is detrimental as it directly impairs an individual's ability to reproduce and pass on genes. But does aging have a benefit at the species level? A benefit emerges when we realize that evolution acting on the individual versus the group is similar to evolution acting on a short versus long time scale. This is logical: traits that benefit an individual are experienced immediately (and operate on a short time scale) while the traits that benefit the species require more time for the benefit to propagate through the group (longer time scale).

As mentioned, on a short time scale, aging and disease are clearly detrimental to an individual's ability to pass on genes. The (longer-term) species-level benefits of shorter lifespans arise from faster turnover in genomes promoting adaption to changes in the environment including competitive encounters with other potentially better-adapted species. From this point of view, the competing individual and species-level pressures make it plausible that a species has an optimal age that is finely tuned. Thus, lifespan would become a *phenotype* (species characteristic) that is tuned to the optimal fitness of the species. We will study how an optimal lifespan striking a balance between these forces might be encoded within a species' genome.

Pressure for shorter lifespans. The forces which strike the balance between individual and group pressures on lifespan are mediated by the environment. Organisms in different environments may have similar genomes but vastly different lifespans. For example, the common rat lives around 2 years while the naked mole rat lives for 30 years (*3*). The naked mole rat spends its life underground and might need more time to live its life - meet mates, have descendants, build community infrastructure, etc. Presumably the optimal time to do this is its lifespan of 30 years (*4*). This suggests that the genome of a specific organism enforces a finite lifespan that is optimal and finely tuned based on the environment and competitive neighboring species. We seek a mathematical model that will, among other things, demonstrate optimal lifespans as finite.

Current research into models for programmed aging includes location-based cellular automata. The locality-based model of Werfel, Ingber, and Bar-Yam includes a simple cellular automaton simulation in which individual grid elements either house a consumer species together with available resources (e.g. food), the resources only, or are empty (*5*). The consumer species members can either die via an intrinsic clock (programmed death at a given age) or after exhausting all resources in their grid element. They model two competing subspecies with different intrinsic mortality rates and thus different average lifespans. The researchers found that while longer-lived individuals exhibited a short-term advantage from longer reproductive spans than their competitors, they also exhibited a long-term disadvantage from local exhaustion of resource supplies from their extended lifespans. Thus, this work demonstrates an advantage for shorter lifespans based on spatial locality and resource depletion. However, there was no investigation into lifespan enforcement mechanisms that are not location-based or the genomic basis for encoding the competition between shorter and longer lifespans.

Initial Dynamical Model

To introduce our genomic dynamics model, let a particular species' genome consist of n genes. For simplicity we assume that each gene i has two main variants labeled $x_i = 0$ or $x_i = 1$. Then generally, a subpopulation within a species has genome $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ where $0 \leq x_i \leq 1$ indicates the proportion of copies of gene i within the subpopulation that have gene variant 1. In this manner, the genome of a subpopulation represents a point $\mathbf{x}$ in an n-dimensional genome space $\mathcal{S} = I^n$ where $I = [0, 1]$ denotes the unit interval. The movement of this point (small subpopulation) in genome space represents the forces of random mutations, evolution, and natural selection changing the proportions of individuals within the subpopulation that have particular gene variants. This yields a time-dependent genome $\mathbf{x}(t)$ for the subpopulation. Figure 1 illustrates this concept in a simple three-dimensional genome space $\mathcal{S}$.

The time evolution of $\mathbf{x}(t)$ involves a stochastic component (due to random mutations) that can be represented as Brownian motion. There is also a drift component due to the evolutionary pressures of the fitness values of the genes. If one variant of a particular gene has a high level of fitness, then we would expect more individuals to have that gene variant in the future. This causes changes to $\mathbf{x}(t)$ and the movement of that point in genome space $\mathcal{S}$.

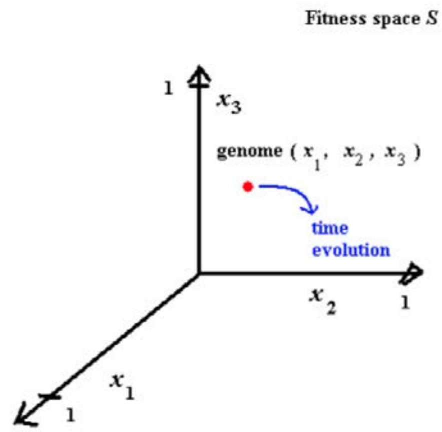


Figure 1: This is an illustration of a three-dimensional genome space $\mathcal{S}$ and the time evolution of a subpopulation genome moving in S.

Zeroth order dynamics. We define $F(\mathbf{x})$ to be the *fitness* of genome $\mathbf{x}$. For example, this can be defined as (a normalization of) the expected number of descendants from $\mathbf{x}$ after some standard time T. We may assume that evolutionary pressure causes the genome $\mathbf{x}$ to drift so as to maximize $F(\mathbf{x})$. Upon rescaling $F(\mathbf{x})$ (by composition with a monotonic function), we may assume that the genomic drift is in the direction of the gradient $\nabla F(\mathbf{x})$ with a speed proportional to this gradient. We can thus describe the time evolution of $\mathbf{x}(t)$ as a dynamical system:

$$\frac{d\mathbf{x}}{dt} = k\nabla F(\mathbf{x}) + \sigma w(t). \tag{1}$$

Here the term $w(t)$ is standard white noise (the derivative of Brownian motion), so that when $k = 0$ and $\sigma = 1$, $\mathbf{x}(t)$ is a Brownian motion process. This then represents diffusive effects of random mutations. On the other hand, $\sigma = 0$ yields a non-stochastic dynamical hill-climbing evolution process aiming for a local maximum of the fitness function $F(\mathbf{x})$.

For simplicity we will suppress dimensions of the genome space $\mathcal{S}$ that have to do with lifespan and consider simply two subpopulations, U_1 and U_2, both with the same initial genome, $\mathbf{x}_1(0) = \mathbf{x}_2(0) = \mathbf{x}_0 = (x_1, \ldots, x_n)$. However, the lifespan (and reproductive span) of U_1 is assumed to be longer than that of U_2. The subpopulation U_1 with its longer lifespan would have an initial evolutionary advantage with larger numbers of descendants for each member of U_1. However, after a longer time, disadvantages manifest for the longer lifespan of U_1 in that this subpopulation will not be able to adapt as fast as the subpopulation U_2 to a changing environment. We model the changing environment by adding a time dependence to the fitness function now in the form $F(\mathbf{x}, t)$. Thus, the hill-climbing process in (Eq. 1) becomes dynamic in the sense that the 'hill' can move. The more frequent generational replenishment of the genome of U_2 with a new one allows the genome $\mathbf{x}_2$ of U_2 to keep up with the local maxima of F.

More specifically, the inter-generational time $m\tau$, which we assume to be proportional to the mean lifespan τ, introduces a fixed time delay $m\tau$ so that in (Eq. 1), $F(\mathbf{x})$ now has the form $F(\mathbf{x}, t + m\tau)$. This time delay makes it impossible for the genome $\mathbf{x}(t)$ to adapt to the current location of the maxima of $F(\mathbf{x}, t+m\tau)$ and thus consistently leaves the genome $\mathbf{x}(t)$ in a region of low fitness with consequent low reproduction rates related to this. We will discuss this model in detail using the diffusion partial differential equation corresponding to the stochastic process (Eq. 1).

Limitations of the zeroth order model. The above simple dynamical model of the time evolution of a subpopulation's genome has some descriptive limitations. In particular, this model can predict very episodic and highly variable epochs in the above dynamical system. For example, let us assume that stable dynamics have arisen through (Eq. 1). This then describes an altruistic state of the subpopulation U_2 in which the shorter lifespan τ dominates. It would then be an advantage for any single individual in this subpopulation to 'cheat' and develop a unilaterally longer lifespan. The locally higher reproduction rate resulting from this would lead to the emergence of a subpopulation with a longer lifespan. As mentioned, this subpopulation would have an immediate (short time scale) advantage by outliving and out-reproducing the neighboring subpopulations. Such longer-lived mutants would quickly become a majority within the family of subpopulations. That is, a new longer-lived family (again denoted as U_1) would again begin to dominate the genomic neighborhood.

However, by the longer-term dynamics described above, the evolutionary forces of (Eq. 1) will dominate in the long run highlighting the adaptive disadvantages of a longer lifespan and again extinguishing the 'cheating' subpopulation U_1. This implies that a full model would involve 'flare-ups' of longer-lived cheaters (subpopulations generically designated as U_1) followed by longer-term extinctions of such cheating subpopulations and eventual re-domination by shorter-lived subpopulations U_2. Such cycles would occur ad infinitum in model (Eq. 1). The fact that such cyclic flare-ups of cheaters in generally altruistic (shorter-lived) populations are seldom observed would indicate that the dynamics of (Eq. 1) are not sufficient to describe global properties of lifespan evolution.

First Order Dynamics

The unstable aspects of the above zeroth order model are not observed in nature; we do not observe cyclic intense flare-ups of longer-living individuals within populations. One might conclude that the mechanisms enforcing optimal lifespan must be so deeply embedded within the genome that simple mutations leading to the emergence of longer-living mutants rarely occur. Alternatively, there might be an enforced 'grammar' in the genome that cannot change rapidly and that only allows certain families of mutations to occur (and fix) easily. If such a grammar were to exist, it would be based on deeper and more detailed dynamics of time evolution in the genome space $\mathcal{S}$. Such dynamics would restrict access to genomic configurations or manifolds

$\mathcal{M} \subset \mathcal{S}$ characterized by stable longevity (denoted as *longevity manifolds*). However, the only restrictions possible in dynamics of a subpopulation U in the genome space $\mathcal{S}$ are based on short survival times there. For this reason, it might be hypothesized that such deeper dynamics block accessibility to longevity manifolds $\mathcal{M}$ by programming 'lethal' genomes $\mathbf{x}_L$ along canonical paths starting in $\mathcal{S}$ and leading into $\mathcal{M}$.

This then begs the question of why a subpopulation cannot 'cheat' its way out of a genomic first order meta-program such as the above, i.e., to evolutionarily remove these higher order block configurations. The point is that it may be possible to program pathways in the above dynamics that make such meta-transitions very slow via even higher order and deeper programmed dynamics.

Purely mathematically, we are discussing a dynamical system (Eq. 1) in which we have expanded the dimensionality to include genes x_i that determine longevity itself so that evolution of longevity can be framed within this dynamical system. Longevity manifolds $\mathcal{M} \subset \mathcal{S}$ would be blocked by dynamics mentioned above in which lethal gene combinations would occur on canonical paths from $\mathcal{S}$ into $\mathcal{M}$. Evolving longevity out of the above configuration would involve changing the dynamics themselves, i.e., changing the very parameters of the genomic dynamical system. Such 'dynamics of dynamics' have been studied in the context of adiabatic evolution of dynamical systems (*6*).

To state things more generically, a more accurate model of the time evolution of a population's genome would feature an 'evolution of evolution' with the process of evolution itself evolving to be more efficient in eliminating the possibility of longer-lived mutants (*7*). This more precise first order model encompasses the above notion of a 'grammar of evolution' along with potentially higher order models forming grammars for lower order ones.

Details of first order dynamics. To posit a first order model is one thing, but it is useful to form some specific examples of such models of genomic dynamics (or meta-dynamics). One general approach to the study of dynamical mechanisms enforcing longer lifespans involves first the introduction of a subspace T of the genome $\mathcal{S}$ that directly affects longevity. If we assume initially that T is one-dimensional, i.e., that there is a single gene determining programmed aging, then we have a deficient model since it is clear that any organism can easily 'cheat' by changing one gene, which can happen easily in one generation. From this, one would hypothesize that the subspace T must be multi-dimensional so that the dynamics within T can be complicated enough to force slow evolution of lifespan.

In a first order dynamical model, we will denote such a multi-dimensional subspace $T \subset \mathcal{S}$ as a *meta-gene* since we are only interested in one phenotype it controls, that being programmed lifespan. We will assume for simplicity that the subspace T contains the first m coordinates of $\mathcal{S}$, so that a genomic location $\mathbf{x}_T \in T$ can be represented as $\mathbf{x}_T = (x_1, \ldots, x_m)$. Further, here we assume that the two far corners of the T 'cube' $\mathbf{0} \equiv (0, 0, \ldots, 0)$ and $\mathbf{1} \equiv (1, 1, \ldots, 1)$ correspond to viable long lifespan and short lifespan genomes, respectively. The challenge to the subpopulation seeking to 'cheat' by traversing from $\mathbf{1}$ to $\mathbf{0}$ is to find an evolutionary path p that does not have any lethal gene combinations $\mathbf{x}_L$ along it.

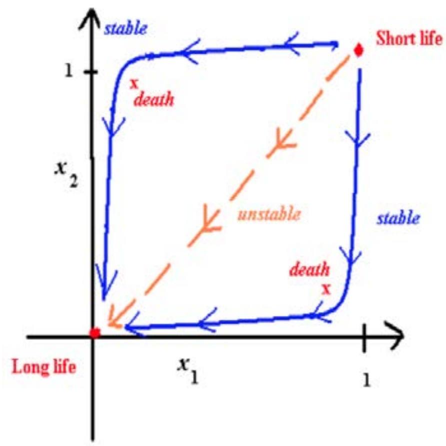


Figure 2: This is a two-dimensional meta-gene for lifespan that illustrates the dynamics of a first order model of evolution.

Consider the simple case of a two-dimensional meta-gene T parameterized by the first two coordinates (x_1, x_2) that control lifespan. This two-dimensional subspace of the entire gene space is illustrated in Figure 2. This figure demonstrates two-dimensional dynamics in a first order model that prevents the transition from a short lifespan $(1, 1)$ to a long one $(0, 0)$. Thus, the meta-gene value $(x_1, x_2) = (0, 0)$ encodes the longer lifespan phenotype, and $(x_1, x_2) = (1, 1)$ encodes the shorter lifespan phenotype. One path for mutating from the short lifespan to the long one is via mutation of both x_1 and x_2 simultaneously as illustrated by the direct diagonal orange path in Figure 2. While this evolution path to a longer lifespan is quicker, it is unstable in requiring a simultaneous mutational 'flip' of the two genes x_1 and x_2, which is probabilistically prohibitive. The alternative and much more likely ('stable') paths to evolving a longer lifespan involve x_1 and x_2 mutating independently and sequentially via the longer blue paths. Within these stable genomic dynamics, the more stable blue paths are blocked because the gene combinations $(0, 1)$ and $(1, 0)$ are assumed to be lethal or have very low survival value (i.e., they have fitness F near 0).

Such a model for two-gene combinations can be extended to three-gene combinations. For example, in such a model, mutations from any single or pair of a three-gene combination could be lethal, but the combination of all three would not be maladaptive and encode a longer lifespan. Such three-gene combinations may occur in nature in maintaining older established diseases (so that they are difficult to evolve out of). We can of course extend such models to n-gene interactions, which become successively harder to defeat through evolutionary adaptation as n grows.

We will extend the mathematical foundation for the dynamical systems approach with a zeroth order model based on (Eq. 1) in the *A PDE Analysis Model* section.

The Role of Disease Genomes in Lifespan Dynamics

It is likely that for any first order dynamics encoding limited lifespan, the subsets or sub-manifolds of genome space $\mathcal{S}$ corresponding to recognized (or unrecognized) disease signatures may play an important role. We will assume here what has been argued above (as well as by others) that evolution tunes the genome to enforce a limited lifespan of a certain length that is optimal for a given species or sub-species. We will consider what role such disease manifolds $\mathcal{M}_1, \mathcal{M}_2, \ldots$ might play mechanistically in the evolution of enforced aging dynamics.

High genomic dimensionality of disease signatures. First we will argue that such disease manifolds $\mathcal{M}$, to the extent they encode 'standard' diseases, should be very high-dimensional. By a standard disease we mean one that has a broad footprint across species and has a long phylogenetic history. Note that if a phylogenetic tree of species susceptible to a certain cancer were to be drawn, some cancer subtypes that are common across all mammals (including humans and mice) must out of necessity also have a long phylogenetic history. That is, they were present in the common ancestors of mice and humans. We will denote such diseases as *standard* or *deep* diseases.

First we note that many deep diseases (e.g. atherosclerosis, diabetes, Alzheimer's) have eluded predictive fingerprinting in genome-wide association (GWAS) and single nucleotide polymorphism (SNP) studies since the completion of the human genome project in the year 2000. Indeed, a sense of disappointment has existed to the extent that 'smoking guns', i.e., clear predictive genomic signatures for these diseases, seem to be missing (*8*). This in itself speaks to the conclusion that the manifolds $\mathcal{M}$ corresponding to the above-mentioned deep diseases must involve many genes, i.e., be high-dimensional.

Such disappointment might have been predicted given the fact that deep diseases out of necessity must have deep high-dimensional signatures that are difficult to predict from. Indeed, if these manifolds were low-dimensional and involved very few genes, then these deep diseases would have been eliminated by evolution by now (i.e., only simple recombinations of a few genes would have been needed to evolve out these diseases). For this reason, it is to be expected that deeper and phylogenetically older diseases should have complicated signatures and thus ones that can be unpredictive.

More specifically, our predictive hypothesis is that older diseases (measured by breadth across species) will be more deeply embedded in the genome (i.e., have complex genomic signatures). A short proof is that a simple genomic signature for disease H means that H can be eliminated from the genome within a few generations. Thus, only complex disease signatures remain. We expect that the most common diseases of old age (e.g. atherosclerosis, cancer, diabetes) should be difficult to predict using GWAS and SNP studies.

The complex embedding of such deep diseases in the genome is necessary to preserve their roles in enforcing lifespans. Future work will involve specific investigations into these ideas, and statistical validation studies of this hypothesis will be conducted.

Evolution and phylogeny of diseases. Given the claim that shorter lifespans are functional evolutionarily, the same claims should be considered for diseases, which can be viewed as enforcement mechanisms for limited lifespans. Mathematically, this would imply that the evolutionary dynamics in genome space $\mathcal{S}$ program a limited lifespan primarily through the manifolds $\mathcal{M}$ that encode diseases. For example, if the escape rate ϵ from a disease manifold $\mathcal{M}$ is sufficiently small, then weaker species-level mechanisms shortening lifespan (i.e., attracting $\mathbf{x}$ back toward $\mathcal{M}$) would be effective.

A model example. Production and activation of high-density lipoproteins (HDL) in humans, for example, could proceed along the above lines. Consider a hypothetical pair of genes without individual main effects (or negative individual effects) interacting to make viable a pathway which produces active HDL and prevents atherosclerosis. In modeling this we can calculate rates of simultaneous two-gene drifts toward longer lifespan (via HDL production) as compared to one-gene drifts which have no effect or a negative effect. Such simultaneous two-gene drifts occur much more slowly (less frequently). Within such regions of slow drift toward longer lifespan, weaker species-level pressures on the species genome $\mathbf{x}$ can compete toward maintaining a shorter lifespan.

Given this interpretation that diseases can play functional roles for species (in terms of lifespan enforcement), they can be treated similarly to other functional phenotypes of an organism, which evolve directionally along its phylogenetic tree. The fact that the functionality of these diseases enforces lifespan may make them as functional and evolutionarily directed as say the evolution of any other phenotype, e.g. the eye or the immune system.

Evolution and cancer. In studying the phylogeny of cancer subtypes, Loso and Tautz discuss the root of the cancer phylogenetic tree being around 600 million years ago (*9*). Here cancer played an initial role in population dynamics of multicellular proto-organisms. It is well-known that individual cell-level apoptosis mechanisms occur both in cells of multicellular organisms and single-celled organisms, e.g. E. coli.

In the case of cell-level apoptosis, there is a bifurcation in a cell's development in which damaged DNA is repaired, or the cell terminates itself. At the root of the phylogenetic tree of multicellular organisms (for organisms with say 10 to 10^4 cells), it is likely that organism-level apoptotis mechanisms came into being in the same way: as adaptive mechanisms to eliminate damaged DNA. This would be an organism-level rather than cell-level mechanism in which sufficient damage to DNA leads to apoptosis at the organism level via cancer initiation. This would allow the initiation of destructive proliferation of damaged cells with the outcome of elimination of the entire organism. Thus, it may be presumed that apoptosis mechanisms that eliminate cells with sufficient DNA damage (e.g. primitive equivalents of the anti-cancer gene P53, which can initiate apoptosis) would have analogs at the species level involving apoptosis of entire organisms.

Multicellular organisms with severely damaged DNA would be sufficiently compromised so that destructive programs (primitive analogs of cancer) would serve to organize apoptosis of the entire organism. This can lead to the view that cancer may serve as a species-level apoptotis tool with mechanisms that are very well developed over an evolutionary lifespan of 600 million years. The study of the phylogeny of cancer as well as other diseases may well be an important area of investigation that we plan to pursue in further work.

A PDE Analysis Model

In this section, we describe a partial differential equation (PDE) model of the time evolution of a population's distribution $\rho(\mathbf{x}, t)$ in genome space $\mathcal{S}$ and how it models an optimal lifespan. In the following *Results* section, we will describe some simulation results.

First we will state some simplified axioms for our model:

- **A1.** The genome of a species encodes an optimal average lifespan.
- **A2.** The lifespan is optimal given the environment of the species.
- **A3.** A changing environment is equivalent to a change in the fitness of the genome.
- **A4.** Lifespan must allow an age of maturation.
- **A5.** Reproduction is asexual via replication of a single parent organism.

Axioms A1 and A2 are the bedrock for this entire analysis as we assume that lifespan is genetically programmed and environmentally optimal. If we take these as given, then a successful model and subsequent simulation will allow us to study the mechanism behind the process. Axiom A3 allows us to model a realistic changing environment as a fitness function, $F(\mathbf{x}, t)$, of the genomic state vector $\mathbf{x}$ and time t. Axiom A4 places a minimum requirement on a genetically programmed lifespan; this realistically should be at least the age to reproduction. Therefore, if τ is the mean lifespan for the population, then we can set $m\tau$ as the time to reproduction with the multiplicative factor $m \in (0, 1)$. Axiom A5 is intended to simplify the modeling of reproduction.

The model is based on the time evolution of the population probability density, which we denote as $\rho(\mathbf{x}, t)$. This is the population density of organisms at the genome (location in genome space) $\mathbf{x} \in \mathcal{S}$ at time t where $\mathcal{S}$ is genome space. This model is the diffusion equation analog of the stochastic dynamical system described in (Eq. 1). The time evolution of the population genome density $\rho(\mathbf{x}, t)$ occurs from two processes. The first is random genetic mutations that can be modeled as a diffusion process in genome space:

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} = c\Delta\rho(\mathbf{x}, t). \tag{2}$$

Here $\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_n^2}$ is the Laplacian operator and c is a diffusion parameter that reflects the DNA mutation rate. If we think of the genome of an individual as a point $\mathbf{x}$ in the gene space $\mathcal{S}$ and follow its movement in $\mathcal{S}$ via the genetic mutations in its DNA germ line, then the above diffusion equation dictates the density ρ as the density of a Brownian motion. This holds if the only time dependence arises from random genetic drift.

The second (evolutionarily directed) process acting on $\rho(\mathbf{x}, t)$ arises from the fitness $F(\mathbf{x}, t)$ of individual genomes $\mathbf{x}$ at time t. Since fitness $F(\mathbf{x}, t)$ is measured by numbers of descendants, it is reasonable to assume that change in $\rho(\mathbf{x}, t)$ is proportional to $F(\mathbf{x}, t)$. This is dictated by the equation

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} = kF(\mathbf{x}, t) \tag{3}$$

where k is a constant. Combining (Eq. 2) and (Eq. 3), we have the following PDE model for the time evolution of the population density in genome space:

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} = c\Delta\rho(\mathbf{x}, t) + kF(\mathbf{x}, t). \tag{4}$$

How should we expect a solution of this equation to behave? If we assume for simplicity that the fitness function is a simple unimodal function on genome space, e.g. a multivariate Gaussian whose mean value changes with time, then we would expect the population genome density $\rho(\mathbf{x}, t)$ to follow $F(\mathbf{x}, t)$. Suppose at a particular genome $\mathbf{x} \in \mathcal{S}$ at time t there is significant overlap between the population density $\rho(\mathbf{x}, t)$ and the fitness function $F(\mathbf{x}, t)$ (i.e., they are both large there). Since $\mathbf{x}$ has high fitness, the genome $\mathbf{x}$ will have multiple offspring resulting in an increase in $\rho(\mathbf{x}, t)$ at the location $\mathbf{x}$. Thus, at the next time step the population density ρ will shift so that its 'peak' or maximum density will be closer to that of the fitness function F. However, since the peak of F is itself moving to reflect a changing environment, the movement of the mass of ρ should mimic the movement of F in gene space.

So how does programmed lifespan fit into these zeroth order dynamics? We first need to consider the effect of a maturation age $m\tau$ as a fraction m of the lifespan τ and how it affects the PDE in (Eq. 4). The first term on the right of (Eq. 4) dictates continuous random genetic mutations leading to a pure diffusion process, i.e., density of a Brownian motion. On the other hand, since the fitness function affects the population density via the number of successful offspring produced, the shift in $\rho(\mathbf{x}, t)$ due to $F(\mathbf{x}, t)$ is not realized until individuals reach maturation age. This results in a necessary time delay of $m\tau$ modifying (Eq. 4) into the following:

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} = c\Delta\rho(\mathbf{x}, t) + kF(\mathbf{x}, t + m\tau). \tag{5}$$

Now we begin to see how a shorter programmed lifespan should be advantageous in this model. For fixed m let the peak of F move away from ρ, so the population density ρ needs to evolve to keep up with the peak of F in order to sustain the population. What dictates how effectively the mass of the population density follows the peak of the fitness function? Since the influence of F on ρ is not realized for a period consisting of the maturation time $m\tau$, this appears as the time delay term in (Eq. 5) above. If lifespan τ is long, then during the length of time it takes for the mass of ρ to shift closer to the peak of F (via successful reproductions), the peak of F itself has moved a considerable amount away from the bulk of ρ. Thus, for long lifespan τ, F can escape from the bulk of the population density ρ which now occurs at very low gene fitness values; as a result the population can collapse. This suggests the mechanism by which an optimal lifespan is favored: the population needs to evolve in order to respond to a changing environment, and the vehicle for this is continuous turnover of the population via short lifespan τ.

It is important to note here that while the above discussion supports our argument for a *short* programmed lifespan, it does not fully extend to an argument for an *optimal* programmed lifespan. To complete the picture, we would need to alter the model based on (Eq. 5) so that there is some penalty for too short a lifespan. An appropriate penalty would be multiplicative reduction factors of the fitness function F. Such a fitness penalty for too short a lifespan is biologically sound; if an organism dies before it matures, for example, then this has a clearly negative impact on the fitness function. In future work, we will implement a biologically sound penalty for lifespans that are shorter than the maturation time of a species resulting in reduced overall fitness.

Results

In this section we discuss results from simulations of the following three scenarios based on the PDE model of (Eq. 5):

1. **Stationary Fitness Function:** The function $F(\mathbf{x}, t) = F(\mathbf{x})$ does not change with time. Initially at $t = 0$, there is little overlap between $\rho(\mathbf{x}, 0)$ and $F(\mathbf{x})$ in gene space.

2. **Moving Fitness Function:** The function $F(\mathbf{x}, t)$ changes with time via a moving center mean.

3. **Moving Fitness Function with Two Subpopulations:** Here we have the same $F(\mathbf{x}, t)$ from simulation #2, but now there are two subpopulations with distinct programmed lifespans τ_1 and τ_2.

Simulation #1 is intended as a 'base' scenario to ensure the efficacy of the simulation and model in general. Here, we hope to see the population gene density $\rho(\mathbf{x}, t)$ evolve to have its bulk mass move toward the peak of the fitness function $F(\mathbf{x})$ in gene space, which itself remains stationary as time progresses. After a sufficiently long simulation time T, we expect to see $\rho(\mathbf{x}, T)$ completely coincide with $F(\mathbf{x})$ if the lifespan is appropriate. Simulation #2 is a straightforward simulation of (Eq. 5); $F(\mathbf{x}, t)$ changes location within gene space with t, and we expect

$\rho(\mathbf{x}, t)$ to evolve to follow it if the lifespan is sufficiently short. Finally, simulation #3 is intended to demonstrate the result of the previous discussion: a shorter programmed lifespan results in greater sustainability of the population size in the face of environment-induced changes to fitness. For this last simulation, half of the starting population is subject to lifespan τ_1 while the other half is subject to lifespan τ_2. If $\tau_1 < \tau_2$, we expect $\int \rho(\mathbf{x}, T) d\mathbf{x}_{\tau_1} > \int \rho(\mathbf{x}, T) d\mathbf{x}_{\tau_2}$. That is, the subpopulation with the shorter lifespan will make up more of the entire population after sufficiently long time T.

As for some specifics of the simulation, we simplify gene space computationally by making it continuous and one dimensional $\mathbb{R}^1$. The gene vector value $\mathbf{x}$ for an individual is just a single real number $\mathbf{x} = \{x_1\} \in \mathbb{R}^1$ rather than a binary 0 or 1 since these can then become any value in (0,1) when averaged over a small local subpopulation. The initial population gene distribution $\rho(\mathbf{x}, 0)$ is Gaussian while the fitness function $F(\mathbf{x}, t)$ is also Gaussian but with a mean center that is to the right of the mean of $\rho(\mathbf{x}, 0)$ on the real line. For simulation #1, the fitness function will remain in this location to the right of the population density in gene space for the entire simulation run. For simulations #2 and #3, the mean of the fitness function will increase by 1 unit per time step so that $F(\mathbf{x}, t)$ moves to the right in gene space. As for directly simulating the PDE in (Eq. 5), we can do this easily by alternating between the two sources of change in $\rho(\mathbf{x}, t)$. For change due to random genetic mutation, we simulate a usual Brownian motion diffusion process with diffusion parameter c by drifting every individual's gene value $\mathbf{x}$ by a random amount $c \cdot N(0, 1)$ at every time step. For change due to fitness, if an individual is at reproduction age $m\tau$ at time t, then it will bud $k \cdot F(\mathbf{x}, t)$ genetically identical offspring. Finally, if an individual is at lifespan age τ, then they are removed from the population.

Figures 3 (a)-(d) show results from simulation #1. The initial population's gene values are sampled from a Gaussian distribution with mean 0, and the stationary fitness function is a Gaussian with mean 10. We implement a short lifespan of $\tau = 2$ time steps and maturation age of $m\tau = 1$; so reproduction age is half of lifespan, i.e., $m = 0.5$. Figures 3 (a), (b), (c), and (d) are snapshots of gene space with the population density and fitness function at time t =0, 5, 10, and 20, respectively. We can see that as time progresses, the population gene density shifts towards more overlap with the fitness function as expected. At time $t = 20$, we can even see the population gene density begin to take on the exact distribution shape of the fitness function. We have selected a short enough lifespan to see the population genome density mimic the fitness function.

Figures 4 (a)-(d) show results from simulation #2. Here, the initial population's gene values are still sampled from a Gaussian with mean 0, and the fitness function is initially a Gaussian with mean 10. The difference here is that the fitness function changes with time; we increase the mean of the fitness function by 1 at each time step. Figures 4 (a) and (b) show snapshots of gene space at time $t = 10$ and $t = 20$ for a population with a short lifespan of $\tau = 2$ while Figures 4 (c) and (d) show the same but for a population with a longer lifespan of $\tau = 4$. For both populations, we maintain a reproduction age of half of the lifespan. We can see that for the shorter-lived population, the gene density is able to follow the fitness function as it moves to the right of gene space. The lifespan for this population is sufficiently short to allow

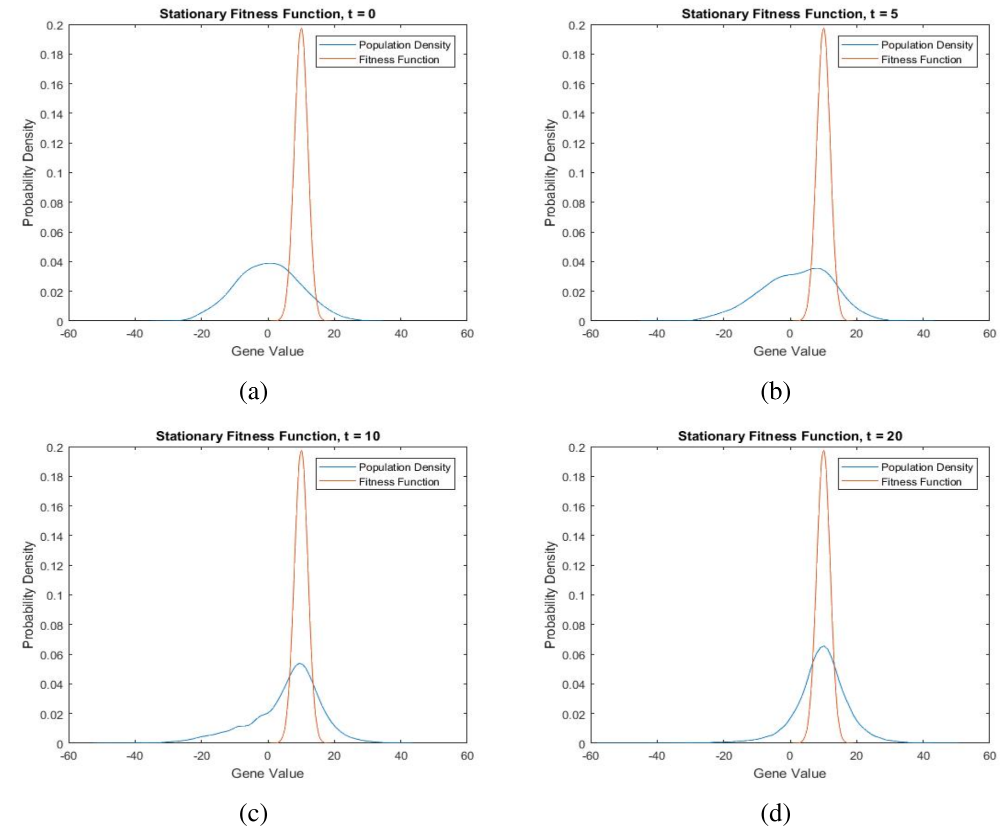
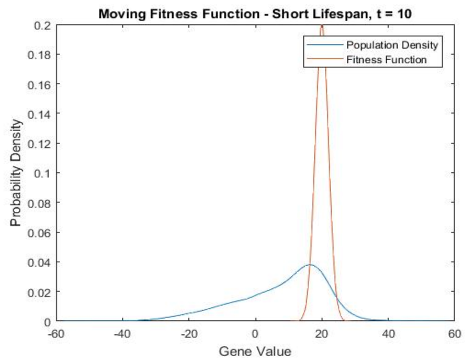


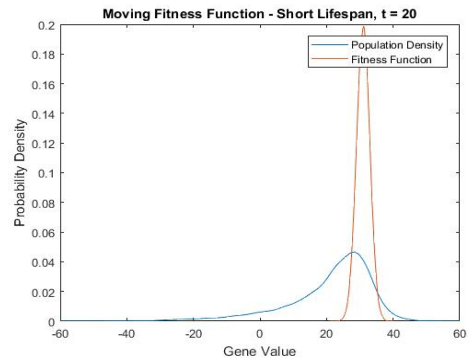
Figure 3: This is the time evolution of a population's genome under the influence of a stationary fitness function at time $t = 0$ (a), 5 (b), 10 (c), and 20 (d). The population has a sufficiently short lifespan of $\tau = 2$ to allow it to mimic the fitness function.

evolution that is rapid enough to respond to the changing environment. On the other hand, for the longer-lived population, the gene density falls too far behind the fitness function, and we see the beginnings of the population's collapse. This population's longer lifespan prevents its members from evolving quickly enough to respond to their changing environment; eventually, the fitness function will have moved so far away from the gene density that the entire population will have extremely low fitness values and reproduce a very small number of offspring.

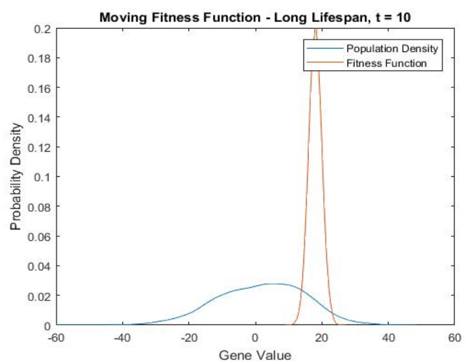
For simulation #3, we still have an initial population gene density sampled from a 0-mean Gaussian and an initial fitness function that is a 10-mean Gaussian whose mean increases by 1 per time step. The difference here is that half of the initial population has a lifespan of $\tau_1 = 2$, and the other half has a lifespan of $\tau_2 = 4$. The analysis from the previous section shows that after a significant amount of time, the shorter-lived subpopulation will be the majority of the entire population. Indeed, after many simulation runs to time $t = 20$, we found that the shorter-lived subpopulation consistently made up approximately 85% of the total population.



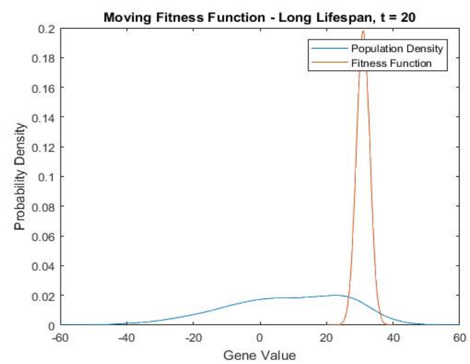
(a)



(b)



(c)



(d)

Figure 4: In (a) and (b), the short lifespan of $\tau = 2$ allows the population to evolve its genome quickly enough to follow the changing fitness function from $t = 10$ to 20. In (c) and (d), the longer lifespan of $\tau = 4$ is too long and results in the population collapsing from $t = 10$ to 20.

Discussion

The evolutionary purpose of aging and disease has been a heavily researched topic for many years that continues to be debated today. In this paper, we describe some models supporting our argument for a species-specific optimal lifespan that is genetically programmed and enforced by disease. The core of the argument is that natural selection operates on a short-term individual scale as well as on a long-term species scale; optimal lifespan is the result of balancing what is good for the individual against what is good for the group. We present a dynamical formalism in the form of a partial differential equation model of population evolution under the influence of random genetic drift and a fitness function. We show analytically as well as via simulation that a shorter programmed lifespan leads to better sustainability of the population in the long run. The reason for this is that a shorter lifespan is necessary for a population to constantly turn over and evolve in response to a changing environment.

We have also discussed the notion that a dynamics of dynamics (equivalently an evolution of evolution) is needed to describe the embedding of programmed lifespans in species' genomes. We have demonstrated embedded dynamics that maintain escape from short lifespan via higher-dimensional evolution. Finally, we have discussed the role that disease manifolds in genome space play in the maintenance of lifespan.

A consequence of these results includes the prediction that 'deeper' phylogenetically embedded diseases will generally have less predictive GWAS signatures. More generally, we have discussed diseases as adaptations toward the maintenance of programmed lifespans. A consequent view involves the evolution of diseases and their phylogenetic trees as important objects of study. In this view, diseases represent adaptive mechanisms that evolve in a directed way like other biological mechanisms such as biochemical pathways and the immune system.

References

1. V. Skulachev, Aging as a particular case of phenoptosis, the programmed death of an organism. *Aging*. **3**, 1120-1123 (2011).

2. M. Nowak, Five rules for the evolution of cooperation. *Science*. **314**, 1560-1563 (2006).

3. E. Kim, X. Fang, A. Fushan, Z. Huang, A. Lobanov, L. Han, S. Marino, X. Sun, A. Turanov, P. Yang, S. Yim, X. Zhao, M. Kasaikina, N. Stoletzki, C. Peng, P. Polak, Z. Xiong, A. Kiezun, Y. Zhu, Y. Chen, G. Kryukov, Q. Zhang, L. Peshkin, L. Yang, R. Bronson, R. Buffenstein, B. Wang, C. Han, Q. Li, L. Chen, W. Zhao, S. Sunyaev, T. Park, G. Zhang, J. Wang, and V. Gladyshev, Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature*. **479**, 223-227 (2011).

4. Y. Zhao, A. Tyshkovskiy, D. Espin-Munoz, X. Tian, M. Serrano, J. de Magalhaes, E. Nevo, V. Gladyshev, A. Seluanov, and V. Gorbunova, Naked mole rats can undergo developmental, oncogene-induced and DNA damage-induced cellular senescence. *PNAS*. **10**, 1073 (2018).

5. J. Werfel, D. Ingber, and Y. Bar-Yam, Programed death is favored by natural selection in spatial systems. *Physical Review Letters*. **114**, 238103 (2015).

6. F. Verhulst, *Nonlinear Differential Equations and Dynamical Systems* (Springer, ed. 2, 2012).

7. J. Mitteldorf and J. Pepper, Senescence as an adaptation to limit the spread of disease. *Journal of Theoretical Biology*. **260**, 186-95 (2009).

8. R. Sherva and L. Farrer, Power and pitfalls of the genome-wide association study approach to identify genes for Alzheimer's disease. *Current Psychiatry Reports*. **13**, 138-146 (2011).

9. T. Loso and D. Tautz, Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC Biology*. **8**, 66 (2010).

Acknowledgments

This research was partially supported by NIGMS Grant 1R01GM131409-01.

Apoptosis and Evolution

Kaczanowski Szymon

Institute of Biochemistry and Biophysics Polish Academy of Sciences,
Pawiskiego 5A, Poland

*To whom correspondence should be addressed; E-mail: szymon@ibb.waw.pl

Important in medicine, apoptosis is a fundamental mechanism involved in multicellular development. We ask the question how mitochondrial apoptosis and metabolism co-evolve. We consider this problem by performing ancestral state reconstruction. Our analysis supports the endosymbiotic theory of apoptosis origin. We found evidence indicating that complexity ancestral apoptotic machinery evolved as a result of an evolutionary arms race between a proto eukaryotic host and a protomitochondrion. According to our reconstruction, the ancestral eubacterial apoptotic machinery contains both caspases and metacaspases, four types of AIFs, both fungal and animal OMI/HTR proteases, and various apoptotic DNases. Different ancient factors were lost in different clades. We hypothesize that acquiring apoptosis and mitochondrial respiration was initially beneficial exclusively in aerobic conditions.This leads to the following expectations: even in extant unicellular eukaryotes, apoptotic factors are involved in mitochondrial respiration, and their activity is beneficial exclusively in aerobic conditions. We test these expectations experimentally using yeasts. Here, we show that known main apoptotic factors deletion are beneficial under anaerobic conditions yet deleterious under aerobic conditions. This observation could have also significant medical implications. It has been shown previously that activity of mitochondrial metabolism correlates with apoptotic activity. Namely it has been shown that inactivation of apoptosis in cancer cells causes frequent respiratory metabolic shifts toward non-mitochondrial glycolysis, known as the Warburg hypothesis of cancer origin. In contrast ,neurons that rely on mitochondrial respiration die due to apoptosis during neurodegenerative diseases (observation known as (Inverse Warburg hypothesis). Indeed there is inverse epidemiological comorbidity between these diseases and cancer.

Introduction

Apoptotic cell death is a fundamental mechanism that regulates multicellular development and is involved in the suppression of cancer. The function of the apoptotic-like cell death observed in unicellular organisms, a form of primordial apoptosis *(1-4)*, is currently unclear. It has been suggested that apoptosis evolved during mitochondrial domestication *(5,6),* as mitochondria are central players in apoptosis and oxidative respiration*(7)*. Some apoptotic factors are vital during mitochondrial respiration4. Pathological inactivation of apoptosis in cancer cells is correlated with frequent respiratory metabolic shifts toward glycolysis (an observation known as the Warburg hypothesis) (*4,8*), and neurons that rely on mitochondrial respiration undergo apoptosis in neurodegenerative diseases such as Alzheimers (*9*) and Parkinsons (*10*) (an observation known as the Inverse Warburg hypothesis) *(4,11)*. Moreover, there is an inverse epidemiological co-morbidity between neurodegenerative diseases and cancer *(12-14)*.

We hypothesized that the observed correlation between mitochondrial metabolism and apoptotic activity resulted from the mitochondrial origin of apoptosis. Using yeast models and phylogenetic analyses, we show that the apoptotic machinery of unicellular organisms is an ancient adaptation to aerobic conditions and respiration.

Utilizing yeast deletion mutants of mitochondrial or cytosolic pro-apoptotic genes as a test of fitness, we found that all mutant strains outperformed wild-type strains under anaerobic conditions and lost their competitive abilities under aerobic conditions. The mutants were not able to grow on medium requiring mitochondrial respiration. Phylogenetic analyses revealed that basic apoptotic mechanisms evolved during mitochondrial domestication.

Additionally, we show that remotely related apoptotic mechanisms described in different clades evolved as a result of the evolutionary race between host and protomitochondria. In ancestral states, apoptotic factors such as different types of proteases, DNases and apoptotic induction factors (AIFs) were toxins, and some were later lost in different clades, such as fungi, which lost caspases. Finally, we also demonstrate that suppressing apoptosis in yeast results in cancer-like remodeling of metabolism.

Experimental evolution

According to the hypothesis formulated by Otto Warburg *(8*), cancer is caused by mitochondrial injury. Subsequent studies showed that suppression of apoptosis is an obligatory element of cancer *(4)* transformation. The endosymbiotic theory of the mitochondrial origin of apoptosis was first proposed by Kroemer, who suggested that apoptotic factors are modified bacterial toxins used by the protomitochondrion during the establishment of symbiosis with protoeukaryotes *(5)*. We hypothesize that apoptosis is an ancient adaptation to aerobic conditions, which require mitochondrial respiration. This hypothesis led to the following predictions: i) assuming that the apoptotic genes of unicellular organisms are involved in aerobic respiration and that their activity is costly, deletion of apoptotic genes should be beneficial under anaerobic conditions;

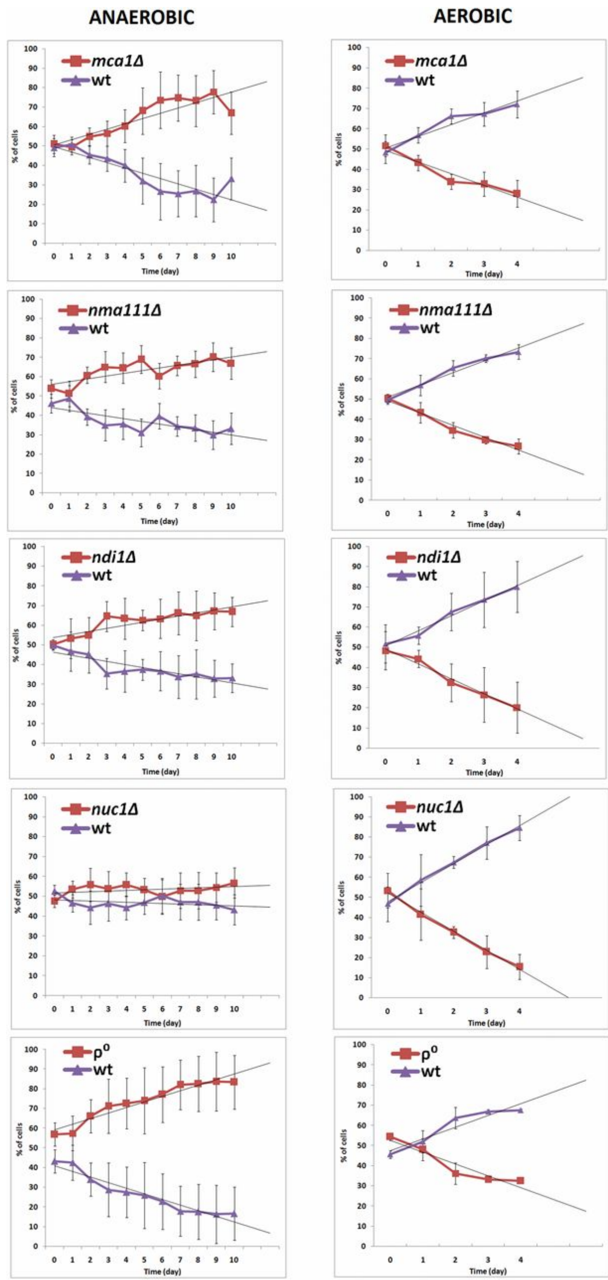


Figure 1: Competition between wild-type and mutant Saccharomyces cerevisiae strains under anaerobic conditions (*19*). Competition assays were performed between ndi1 (A), nuc1 (B), mca1 (C) or nma111 (D) and wild-type S. cerevisiae MR6 strains. The percentage of mutant and wild-type strain cells was monitored every 24 h before sub-culturing into fresh medium (via the plating of approximately 100 colonies onto a rich YPD medium) and replication in YPD medium with geneticin. The data presented in each graph represent the mean SD of three independent assays.

and ii) deletion of apoptotic factors should lead to cancer-like remodeling of metabolism, i.e., suppression of mitochondrial metabolism.

To test these predictions, we used yeast deletion mutants for both mitochondrial (ndi1, nuc1) and cytoplasmic (mca1, nma111) apoptotic factors. Ndi1 is a yeast apoptotic induction factor (AIF) *(15)*, Nuc1 is the EndoG *(16)* apoptotic DNAse, Mca1 *(17)* is a yeast metacaspase, and Nma111 is an apoptotic protease *(18)*. To determine whether these deletions are beneficial under anaerobic conditions, we conducted a competitive experiment to test the proliferation of wild-type and mutant yeast co-cultivated under anaerobic conditions. Equal amounts of cells from the two strains (from the early exponential growth phase) were mixed and repeatedly sub-cultured in O2-free liquid medium. Aliquots of each co-culture passage were plated on YPD, and the number of wild-type and mutant strain colonies was determined (*19)*.

As shown in Figures the mutant cells outperformed the wild-type cells, excluding the experiment assessing the ndi1 and nuc1 yeast mutants; deletion of the apoptotic genes ndil and nuc1 became deleterious under aerobic conditions. Notably, all of the yeast mutant strains showed increased growth under anaerobic conditions, as measured based on optical density, which was not the case under aerobic conditions. Moreover, all mutant strains were unable to grow on a non-fermentable carbon source, glycerol, suggesting that at least some apoptotic proteins are required for mitochondrial respiration. In conclusion, we found that inactivation of apoptosis in unicellular organisms could be beneficial under anaerobic conditions and that the apoptotic machinery is required for cellular respiration. Our results demonstrate that in extant yeast, the apoptotic machinery is an adaptation to aerobic conditions.

Ancestral state reconstruction

We hypothesized that this adaptation evolved during the domestication of mitochondria. This hypothesis is supported by the observation that suppression of yeast apoptosis leads to cancer-like remodeling of metabolism, in which apoptosis is also suppressed. The ancient origin of yeast apoptotic mechanisms leads to the prediction that these mechanisms evolved due to divergent evolution from the ancestral apoptotic machinery. This proposal is not obvious, as the apoptotic machinery of yeast is very different from that of animals and plants. Plants use the ZEN1 apoptotic DNase instead of the yeast DNase ENDOG. Additionally, in animals, metacaspase is replaced by another apoptotic protease referred to as caspase *(4,20)*. These observations prompt the following question: What is the evolutionary origin of such apoptotic mechanisms convergent evolution, divergent evolution, or a combination of the two?

To find a proxy answer to this question, we performed ancestral state reconstruction. We took advantage of recent advances in systematics that revealed six to eight major eukaryotic branches early in evolutionary history. Our results are based on sequencing and genomic data from the following major eukaryotic groups: Opisthokonta (fungi and animals); Amoebozoa (Dictyostelium); SAR (Stramenopiles, Alveolates, and Rhizaria)-(ciliates, apicomplexan parasites and Reticulomyxa); Excavata (kinetoplastids, Trichomonas and Naegleria); and Archae-

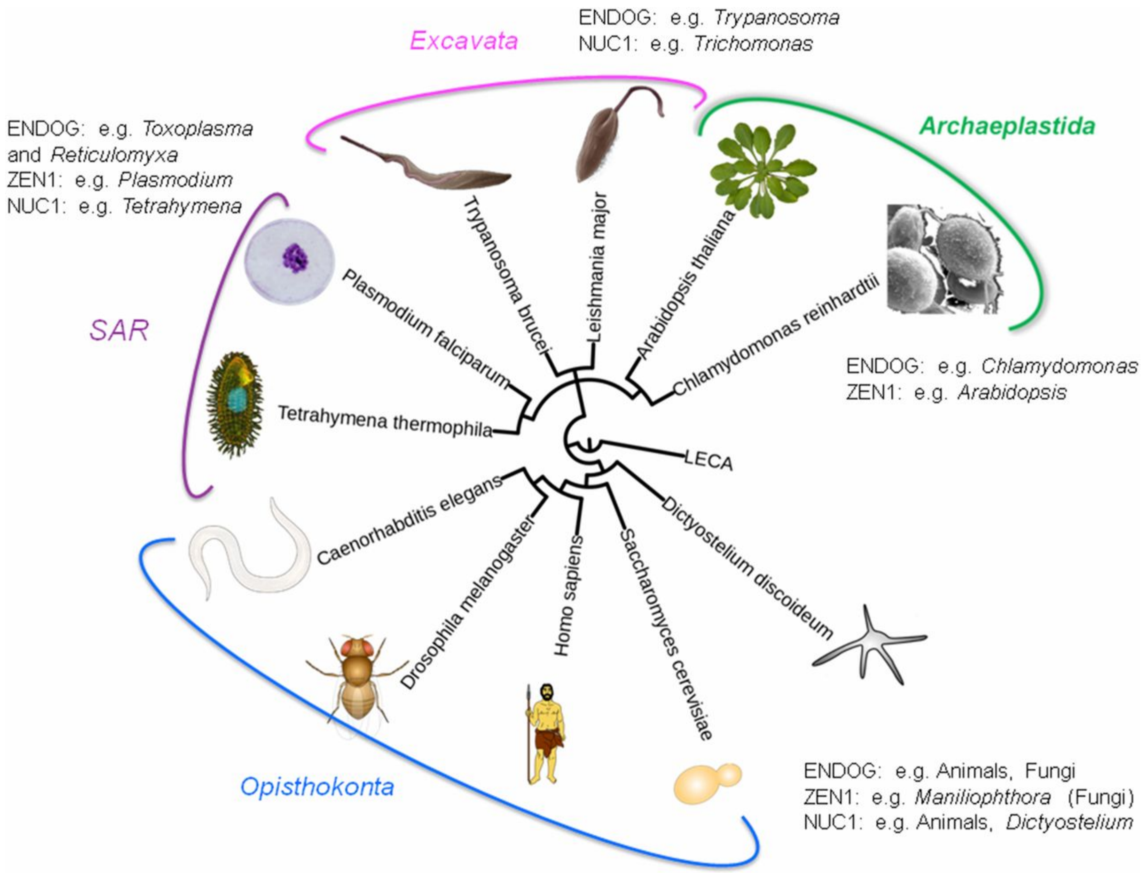


Figure 2: The evolutionary tree of eukaryotes and the evolution of apoptotic DNases (*19*).The applied phylogenetic tree is based on a recently published analysis of the evolution of 37 proteins and shows organisms with described apoptosis (*4, 24*). The authors obtained bootstrap values of 100% for the presented branches

plastida (plants and green algae such as Volvox) *(21)*(see Fig. 2). Using a parsimony assumption, we reasoned that an apoptotic factor would have been part of the ancestral machinery if it exhibited a homolog in organisms belonging to several of these ancient taxonomic groups or in non-eukaryotic organisms. Below, we present examples of the applied analysis, and the full analysis is included in the Supplementary Materials section. Additionally, the basic logic of the analysis is presented for DNases.

The following appear to be ancient DNases: ENDOG, ZEN1, and NUC *(14,20)*. The fact that the apoptotic function of these proteins evolved very early, as this observation suggests, was confirmed by our sequence analysis (Fig. 2). ENDOG proteins contain the Protein Families Database (PFAM) Endonuclease_NS domain (DNA/RNA non-specific domain), which is mainly present in eukaryotic and eubacterial proteins (as well as some archaeal and viral proteins) *(22)*. Phylogenetic trees inferred using maximum likelihood revealed that ENDOG proteins belong to a monophyletic branch that includes eubacterial (but not archaeal) proteins, suggesting that the ENDOG protein has a mitochondrial/eubacterial origin. This hypothesis is supported by the fact that ENDOG is a mitochondrial protein. This branch also contains ENDOG homologs encoded in the genomes of organisms belonging to other major eukaryotic taxonomic groups, namely SAR and Archaeplastida. Thus, we conclude that ENDOG was a component of the protomitochondrial apoptotic system. Similar analyses were performed for the ZEN1 and NUC1 proteins (Fig. 2). A more complex analysis was performed for AIFs and apoptotic proteases. A detailed sequence analysis revealed that AIFs have a eubacterial origin and that many types of AIFs diverged before the origin of eukaryotes. Several pieces of evidence support this conclusion. We calculated a tree using 12,420 sequences containing different AIFs. The different AIFs were often more closely related to bacterial homologs than to other eukaryotic AIFs. We identified at least four putative ancient branches. We further tested this hypothesis using phylogenetic analyses of representative sequences from the three human AIF and NDI branches. In each case, representative bacterial sequences and sequences belonging to each branch were included. The phylogenetic tree calculated using this alignment is shown in Figure 3. There are two old branches, supported by high bootstrap values (94-99%), containing both eukaryotic proteins and the AIFM1/AIFM3 and NDI1/AIFM2 bacterial branches. The chronology of evolutionary events predicted by the RelTime method indicated that the root is placed between these branches. The AIFM1/AIFM3 branch contains a clear sub-AIFM1 sub-branch supported by high bootstrap values (97-99%) and includes protein sequences from the following ancient eukaryotic taxonomic clades: Excavata (Naegleria), Amoebozoa (Dictyostelium), and Opisthokonta (Metazoa).

The NDI1 branch is supported by high bootstrap values (93-99%) and contains both eubacterial proteins and sequences from five major ancient eukaryotic groups (Opisthokonta, SAR, Archaeplastida, Excavata, and Amoebozoa). The AIFM2 branch, with lower bootstrap values of 76-89%, also contains both eubacterial proteins and proteins from five ancient eukaryotic taxonomic groups.

In conclusion, phylogenetic trees based on the alignment of representative sequences indicate that divergence of the AIFM1, AIFM2, AIFM3, and NDI sequences occurred before the divergence of eukaryotes. This conclusion is supported by an analysis of the domain architecture. Both eukaryotic and eubacterial AIFM1 proteins contain the AIF_C domain, and the AIFM3 protein contains the Rieske domain. We also confirmed our observations using BLAST searches. We performed an analogous analysis for apoptotic proteases. In animals, apoptosis is induced by proteases known as caspases. In fungi and plants, a similar function is played by remotely related proteases known as metacaspases. Both proteases inactivate the Tudor nuclease to induce apoptosis. Because plants and fungi are remotely related (plants belong to Archaeplastida and fungi to Opisthokonta), we can assume that metacaspase was part of the ancestral apoptotic machinery. Our detailed sequence analysis suggested that the same is true for caspases. Indeed, it is likely that animal caspases are closely related to putative caspase homologs of Reticulomyxa from the SAR taxonomic group. This homology was supported by different phylogenetic trees with bootstrap values of 90-97% . Analysis of the chronology of

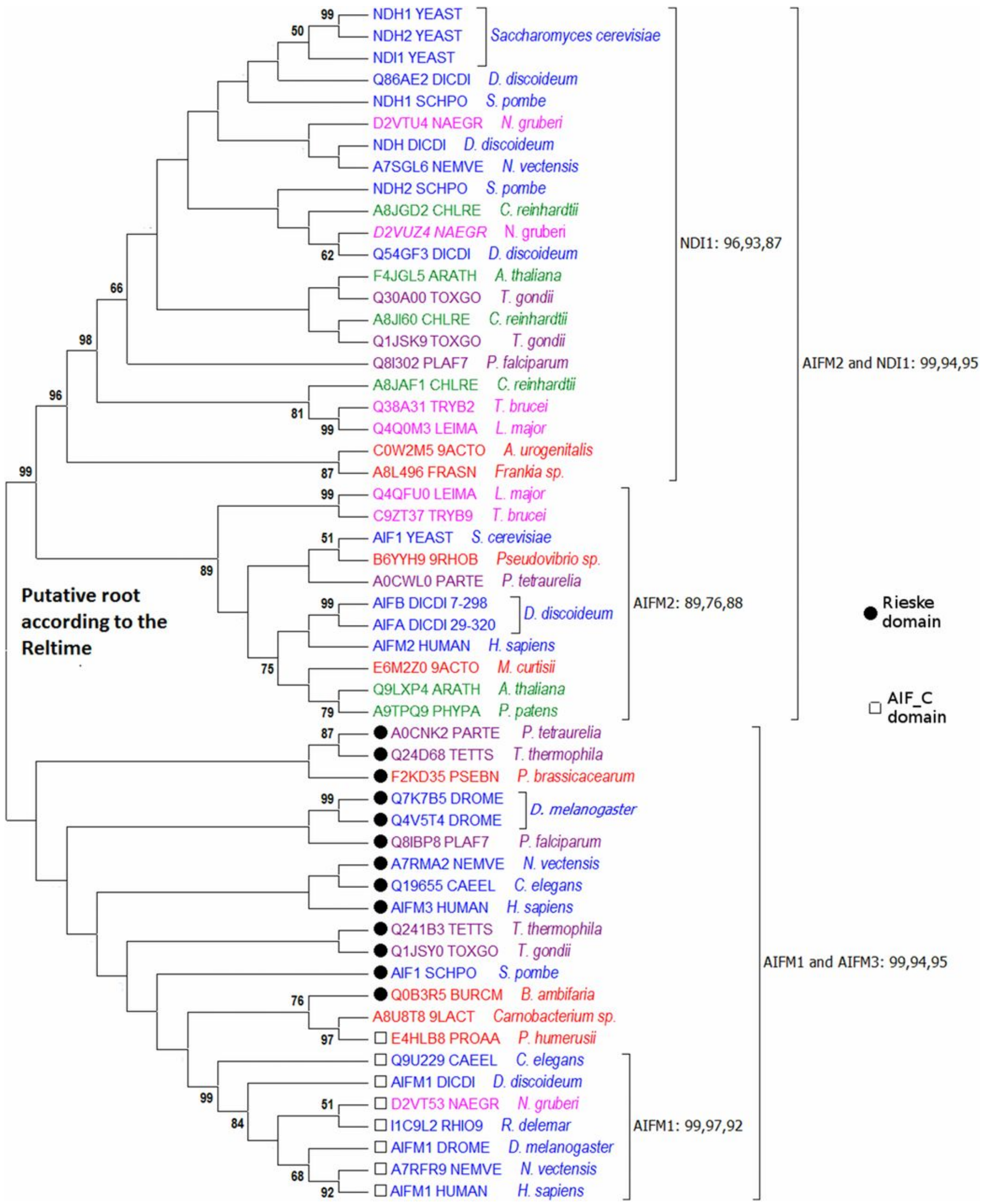


Figure 3: Phylogenetic trees of AIFs. Three comma-separated bootstrap values (100 replicates) for the most significant branches were calculated with maximum likelihood estimation (MLE), neighbor-joining (NJ), and minimal evolution (ME), respectively. The red caption indicates bacterial proteins, black Archaea, dark blue Opisthokonta/Amoebozoa, light blue Archaeplastida, purple SAR proteins, and pink Excavata proteins.

events displayed by the phylogenetic trees did not support the horizontal transfer hypothesis. In conclusion, the results presented above suggest that the protomitochondrion likely encoded both caspases and metacaspases. HTRA proteases are involved in regulated proteolysis and quality control in both eubacterial and eukaryotic organisms. Some of these proteases exhibit an apoptotic function in Opisthokonta, in which they inhibit caspase apoptosis inhibitors containing BIR domains, the so-called inhibitor of apoptosis proteins (IAPs). This inhibition of inhibitors activates apoptosis. Our sequence

analysis revealed that these seemingly typical orthologous proteins diverged before the appearance of eukaryotes, which is supported by bootstrap values of 86-93%. Both human and fungal sub-branches contain plant and bacterial proteases, following phylogenetic order.

As mentioned above, proteins from the OMI/HTRA family induce apoptosis through the degradation of survivin/BIR1p apoptosis inhibitors in both animals and fungi. The BIR domain belongs to the BIR-like clan (PFAM ID CL0417), which includes remotely evolutionarily related protein domains. The proteins belonging to this clan are exclusively viral and eukaryotic and are present in the main eukaryotic taxonomic groups. We conclude that it is likely that BIRs are ancient inhibitors of apoptosis. A similar analysis was performed for the API5-AAC11 inhibitor.

Conclusions

Our analysis suggests that a co-evolutionary arms race contributed to the formation of the complex apoptotic regulatory pathways presented in Figure 4. The surprising richness of redundant apoptotic factors in the protomitochondrion suggests that red queen *(23)* co-evolution may have shaped the protomitochondrion to contain as many toxins as possible . It is difficult for the host to evolve resistance against many similar, but slightly different, pathogenic factors. Experiments performed with yeast models suggest that ancient apoptotic toxins were directly involved in respiration and adaptations to aerobic conditions.

We found that inactivation of apoptosis in unicellular organisms could be beneficial under anaerobic conditions and that the apoptotic machinery is required for cellular respiration. Our results support the hypothesis that the apoptotic machinery has evolved during the domestication of mitochondria and that the impact of pro-apoptotic factors on cellular metabolism is likely a general phenomenon in other eukaryotes, including protists, plants and animals. This hypothesis is supported by classical observations that in the pathological stages of animal cells, the remodeling of metabolism co-occurs with changes in apoptotic activity.

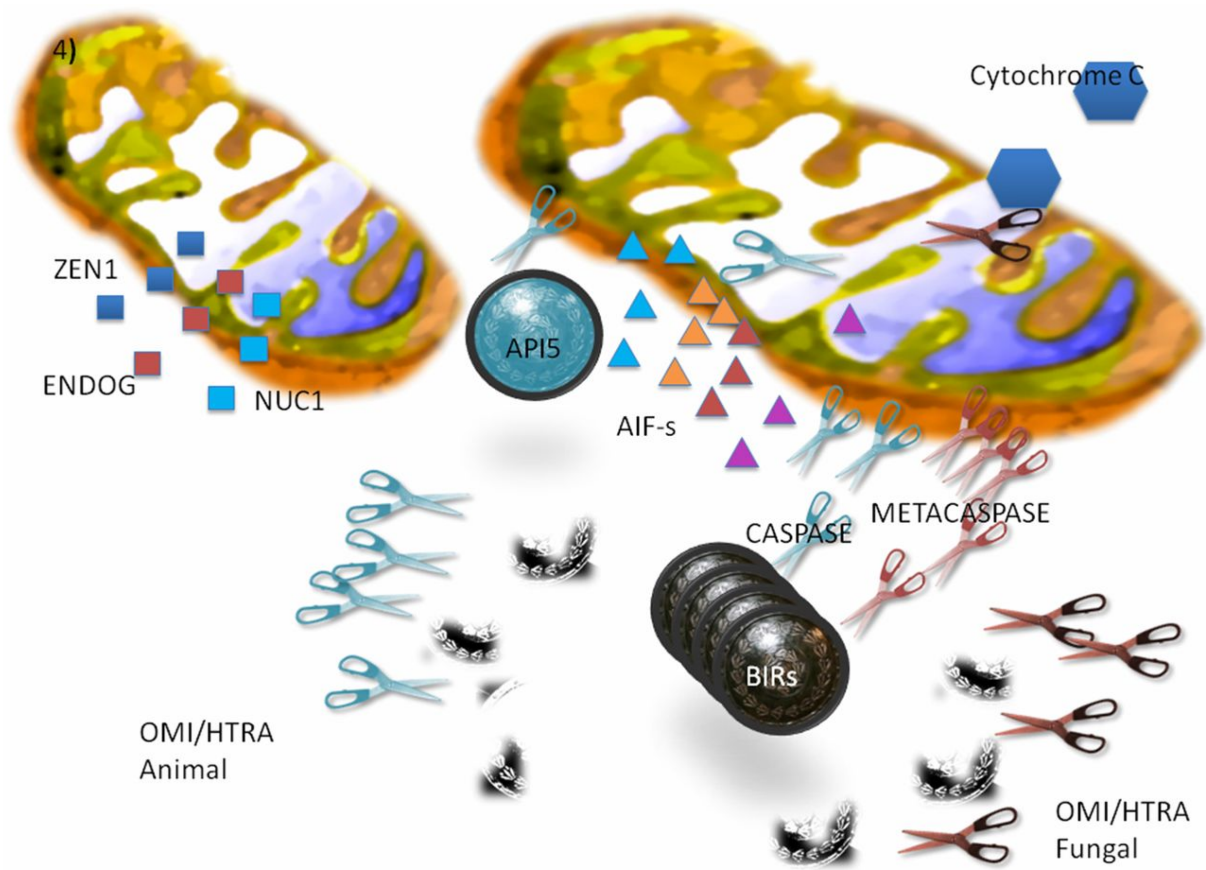


Figure 4: Reconstruction of the ancestral state (*19*). According to this reconstruction, the parasitic protomitochondrion released the following toxins: DNases ZEN1, ENDOG and NUC1 (represented as squares), AIFs (represented as triangles), and various proteases (represented as scissors). Inhibitors of proteases, caspases and metacaspases represent the evolutionary response of protoeukaryotic cells (represented as shields: BIRs in black, API5-AAC11 in blue). The protomitochondrial proteases OMI/HTRA, which deactivate BIRs, are an evolutionary response to BIR. As a result, two similar programs have been established containing metacaspases, caspases, BIRs, and OMI/HTRA proteases. It is likely that the two programs diverged very early in the process of evolution.

References and Notes

1. W. R. Proto, G. H. Coombs, J. C. Mottram, Cell death in parasitic protozoa: regulated or incidental? Nat Rev Microbiol **11**, 58-66 (2013).
2. E. Taylor-Brown, H. Hurd, The first suicides: a legacy inherited by parasitic protozoans from prokaryote ancestors. Parasit Vectors **6**, 108 (2013).
3. P. M. Durand, S. Sym, R. E. Michod, Programmed Cell Death and Complexity

in Microbial Systems. Curr Biol **26**, R587-593 (2016).

4. S. Kaczanowski, Apoptosis: its origin, history, maintenance and the medical implications for cancer and aging. Phys Biol **13**, 031001 (2016).
5. G. Kroemer, Mitochondrial implication in apoptosis. Towards an endosymbiont hypothesis of apoptosis evolution. Cell Death Differ **4**, 443-456 (1997).
6. E. V. Koonin, L. Aravind, Origin and evolution of eukaryotic apoptosis: the bacterial connection. Cell Death Differ **9**, 394-404 (2002).
7. P. Marchetti et al., Mitochondrial permeability transition is a central coordinating event of apoptosis. J Exp Med **184**, 1155-1160 (1996).
8. O. WARBURG, On the origin of cancer cells. Science 123, 309-314 (1956).
9. F. M. LaFerla, B. T. Tinkle, C. J. Bieberich, C. C. Haudenschild, G. Jay, The Alzheimer's A beta peptide induces neurodegeneration and apoptotic cell death in transgenic mice. Nat Genet **9**, 21-30 (1995).
10. H. Mochizuki, K. Goto, H. Mori, Y. Mizuno, Histochemical detection of apoptosis in Parkinson's disease. J Neurol Sci **137**, 120-123 (1996).
11. L. A. Demetrius, P. J. Magistretti, L. Pellerin, Alzheimer's disease: the amyloid hypothesis and the Inverse Warburg effect. Front Physiol 5, 522 (2014).
12. M. Musicco et al., Inverse occurrence of cancer and Alzheimer disease: a population-based incidence study. Neurology **81**, 322-328 (2013).
13. B. Jansson, J. Jankovic, Low cancer rates among patients with Parkinson's disease. Ann Neurol **17**, 505-509 (1985).
14. K. Ibez, C. Boullosa, R. Tabars-Seisdedos, A. Baudot, A. Valencia, Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses. PLoS Genet **10**, e1004173 (2014).
15. W. Li et al., Yeast AMID homologue Ndi1p displays respiration-restricted apoptotic activity and is involved in chronological aging. Mol Biol Cell **17**, 1802-1811 (2006).
16. S. Bttner et al., Endonuclease G regulates budding yeast life and death. Mol Cell 25, 233-246 (2007).
17. R. E. Lee, S. Brunette, L. G. Puente, L. A. Megeney, Metacaspase Yca1 is required for clearance of insoluble protein aggregates. Proc Natl Acad Sci U S A **107**, 13348-13353 (2010).
18. B. Fahrenkrog, Nma111p, the pro-apoptotic HtrA-like nuclear serine protease in Saccharomyces cerevisiae: a short survey. Biochem Soc Trans **39**, 1499-1501 (2011).

19. J. Klim, A. Gadki, R. Kucharczyk, U. Zielenkiewicz, S. Kaczanowski, Ancestral State Reconstruction of the Apoptosis Machinery in the Common Ancestor of Eukaryotes. G3 (Bethesda) **8**, 2121-2134 (2018).
20. S. Kaczanowski, M. Sajid, S. E. Reece, Evolution of apoptosis-like programmed cell death in unicellular protozoan parasites. Parasit Vectors **4**, 44 (2011).
21. S. M. Adl et al., The revised classification of eukaryotes. J Eukaryot Microbiol 59, 429-493 (2012).
22. R. D. Finn et al., The Pfam protein families database. Nucleic Acids Res **38**, D211-222 (2010).
23. L. V. Valen, A new evolutionary law. Evolutionary Theory 1, 1-30 (1973).
24. D. He et al., An alternative root for the eukaryote tree of life. Curr Biol **24**, 465-470 (2014).

Acknowledgments

This research was funded by National Science Centre of Poland NCN grant number 2017/27/B/NZ8/02502.

Influential Quorum Sensing Proteins of Multidrug Resistant *Proteus mirabilis* Causing Urinary Tract Infections

Shrikant Pawar,[1,2] Md. Izhar Ashraf,[3,4] Shama Mujawar[5] Chandrajit Lahiri [5*]

[1]Department of Computer Science, Georgia State University, Atlanta, Georgia, USA - 30303
[2]Department of Biology, Georgia State University, Atlanta, Georgia, USA - 30303
[3]The Institute of Mathematical Sciences, Chennai - 600113, India
[4]B.S. Abdur Rahman Crescent Institute of Science & Technology, Chennai - 600048, India
[5]Department of Biological Sciences, Sunway University, 47500 Petaling Jaya, Selangor, Malaysia

*To whom correspondence should be addressed; E-mail: chandrajitl@sunway.edu.my

Catheter-associated urinary tract infections (CAUTI) has become an alarming hospital based disease with the increase of multidrug resistance (MDR) strains of Proteus mirabilis. High prevalence of long-term hospital based CAUTI for patients along with moderate percentage of morbidity due to ignorance, failure and MDR, necessitates an immediate intervention strategy to combat the deadly disease. Several reports and reviews focus on revealing the important genes and proteins essential to tackle CAUTI caused by P. mirabilis. Despite longitudinal studies and methodical strategies to circumvent the issues, effective means of unearthing the most influential proteins to target for therapeutic uses have been meagre. Here we have reported a strategic approach for identifying the most influential proteins from the complete set of proteins of the whole genome of P. mirabilis, besides comparing the interactomes comprising the autoinducer-2 (AI-2) biosynthetic pathway along with other proteins involved in biofilm formation and responsible for virulence. Essentially, we have adopted a computational network model based approach to construct a set of small protein interaction networks (SPIN) along with the whole genome (GPIN) to identify, albeit theoretically, the most significant proteins. These might actually be responsible for the phenomenon of quorum sensing (QS) and biofilm formation and thus, could be therapeutically targeted to fight out the

MDR threats to antibiotics of P. mirabilis. Our approach signifies the eigenvector centrality coupled with k-core analyses to be a better measure in addressing the pressing issues.

Introduction

Urinary tract infections (UTI) are the second most common infection prevalent amongst long-term hospital patients, second only to pneumonia. Failure to treat or a delay in treatment can result in systemic inflammatory response syndrome (SIRS), which carries a mortality rate of 20-50 percent [1-3]. While Escherichia coli remains the most often implicated cause of UTI in previously healthy outpatients, Proteus mirabilis take the lead for catheter-associated UTI (CAUTI), causing 10-44 percent of long-term CAUTIs [1, 3]. In comparison to normal cases, CAUTI is quite complicated and encountered by patients with multiple prior episodes of UTI, multiple antibiotic treatments, urinary tract obstruction and/or undergoing catheterization as also for those with spinal cord injury or anatomical abnormality [1-3]. Such complications of CAUTI caused by P. mirabilis arise from the usage of a diverse set of virulence factors by the organism to access and colonize the host urinary tract. These include, but not limited to, urease and stone formation, fimbriae and other adhesins, iron and zinc acquisition, proteases and toxins and biofilm formation [1]. Despite significant advances made for studying P. mirabilis pathogenesis, a meagre knowledge of its regulatory mechanism poses an urgent and pressing need to come up with unique health intervention processes for such patients. In attempts to provide such health interventions, longitudinal and epidemiological studies on P. mirabilis have been reported for extended-spectrum -lactamase (ESBL) and AmpC -lactamase .Furthermore, along with various other components of the membrane, several cytoplasmic factors interplay among themselves to regulate the cell-density dependent gene regulation. This enables the bacteria for cell-to-cell communication, a phenomenon known as quorum sensing (QS) [8]. Besides other phenotypic traits, QS controls the expression of the virulence factors responsible for pathogenesis of P. mirabilis [9]. Again, as per other reports, despite producing two cyclic dipeptides and encoding LuxS-dependent quorum sensing molecule, AI-2, during swarming, P. mirabilis has been reported to be have no strong evidence of QS (1,10-12). However, a highly ordered swarm cycle suggests an existing mechanism for multicellular coordination (13). Thus, the fact that MDR P. mirabilis are engaged in biofilm formation which is managed, albeit in parts, through quorum sensing brings out the complexity of CAUTI. To deal with such complexity, analyses of the mosaic mesh or networks of interacting proteins involved, commonly known as protein interaction networks (PINs), can provide sufficient insight to reveal the influential key role players of the phenomenon [14, 15]. The influential role players of phenomenon like QS can be determined by analysing the PIN involving the proteins in the pathway to produce the QS inducer. The essentiality of such small protein interactome (SPIN) can be brought about by an analysis for the most biologically relevant protein to target for inhibiting that phenomenon, also known as quorum quenching. Ideally, a determination of the number of interacting partners of a particular protein identifies its degree centrality (DC) which correlates with its essential nature in the biological scenario [16]. However, a deeper understanding of the essential nature of a particular protein comes upon analysing its interaction with other partners in the global network of all proteins. In this study, we have analysed the importance of other centrality measures like Closeness centrality (CC), Betweenness centrality (BC) and Eigenvector centrality (EC) [16] parameters for SPIN comprising the genes and proteins involved in quorum sensing. Again, analyses of a stipulated sets of QS proteins for a valuable knowledge about the most influential virulence proteins to render as drug targets for the QS phenomenon could be quite insufficient. Thus, we have carried out further analyses of the whole genome of P. mirabilis for a global analysis of the encoding proteins. This comprises the decomposition of the whole genome protein interactome (GPIN) to a core of highly interacting proteins through the k-core analysis approach [17]. Furthermore, to identify the functional modules in the global network [18], we have performed cartographic analyses and predicted the indispensability of certain sets of proteins which have been shown to be sharing similar functional modules empirically important for drug targets.

Materials and Methods

Dataset Collection

The P. mirabilis QS pathways for autoinducer-2 (AI-2) biosynthesis were collected from curated reference databases of genomes and metabolic pathways like KEGG, MetaCyc and BioCyc [19, 20, 21]. The proteins involved in these pathways were extracted and fed as queries to the STRING 10.5 biological meta-database [22] to retrieve protein interaction datasets with at least 10 or 50 interactors having the default medium level confidence [period of access: January to February, 2018]. Detailed protein links file under the accession number 529507 in STRING was used to collect all the interactions of the whole genome proteins of P. mirabilis strain HI4320.

Interactome Construction

We have taken a stepwise approach to integrate and build the interactomes of the proteins. These are the small protein interactomes (SPIN) comprised of a) those involved in AI-2 biosynthetic pathway in the organism with small (10) and large (50) number of interactors (AIPS, AIPL, respectively), b) only QS genes found (QSPO), c) all QS genes reported as homologues (QSPH) present in P. mirabilis, d) all virulent genes reported (QSPV) and e) the whole genome of P. mirabilis (WGPM). The number of P. mirabilis proteins from the SPIN class of interactomes were 31 for AIPS, 102 for AIPL, 24 for QSPO, 42 for QSPH, 58 for QSPV and 3548 for GPIN. The individual protein interaction data, with medium confidence default values, were obtained from String 10.5. Interactions were 80 for AIPS, 435 for AIPL, 30 for QSPO, 129 for QSPH, 426 for QSPV and 358984 for GPIN, respectively. All individual interaction data obtained above were imported into Cytoscape version 3.6.0 [23] and Gephi 0.9.2 [24] to integrate, build and analyse five SPIN namely AIPS, AIPL, QSPO, QSPH and QSPV and the GPIN. Interactomes were considered as undirected graphs represented by G = (V, E) consisting of a finite set of V vertices (or nodes) and E edges. An edge e = (u,v) connects two vertices (nodes) u and v. Each protein is represented as a vertex/node. The number of physical and functional connections/interactions/associations/links a node has with other nodes comprises its degree d (v) [25].

Network Analyses

SPIN

All the five SPIN were subsequently analyzed individually by utilizing the four important concepts of centrality applied to biological networks, namely, degree centrality (DC), closeness centrality (CC), betweenness centrality (BC) and eigenvector centrality (EC) [26-28]. This was done either via Gephi or the Cytoscape integrated java plugin CytoNCA [29]. The combined scores from different parameters considered in STRING were taken as edge weights for computing CytoNCA scores. Top 5 proteins for each of the centrality measures were taken for drawing venn diagrams through online tool Venny 2.1 [30] to find common proteins from each measures.

GPIN

Further analyses for GPIN were done by using MATLAB version 7.11, a programming language developed by MathWorks [31]. A primary understanding of the GPIN was obtained by plotting the distributions of network degree (k) against the Complementary Cumulative Distribution Function (CCDF). An idea of the core group of the very specific proteins was obtained from a K-core analysis of the proteins in the whole genome context. This essentially prunes the network to a k-core with nodes having degree at least equal to k and classifying them in k-shell (proteins, in our study) based on the variety of their interacting partners. This was done following the network decomposition (pruning) techniques to produce a sequence of subgraph of gradually increasing cohesion [17]. Further, a significant knowledge of the functional connectivity and participation of each protein was derived from the cartographic representation of the within-module degree zscore of the protein versus its participation coefficient, P, as per the methodology described by Guimera, et al. [32]. The z-score measures how 'well connected' a node i is to other nodes in the module while P measures how the node i is positioned in its own module and with respect to other modules. Participation of each protein reflected its intra- and inter-modular positioning, where functional modules were calculated based on Rosvall method [33]. Such analyses demanded the proteins to be mainly divided into two major categories namely the hub nodes and the non-hub nodes where the former is a connection point of many nodes. The category of latter has been assigned roles of ultra-peripheral nodes (R1), peripheral nodes (R2), non-hub connector nodes (R3) and the non-hub kinless nodes (R4). Likewise, the hub nodes have been assigned as provincial hubs (R5), connector hubs (R6) and kinless hubs (R7) [32].

Results

The individual five SPIN

To have an understanding of the important protein(s) of QS in P. mirabilis, we have taken a stepwise approach of building five SPIN, with an ultimate goal to identify the influential virulent proteins to serve as potential candidates for therapeutic targets. A comparative picture of the parametric values of the top five rank holders in their descending order has been delineated in a tabular form (Table 1).

Table 1. The top 5 rank holders of P. mirabilis SPIN and GPIN. The bold cased proteins are present in the innermost 154th k- core.

Network	EC	BC	DC	CC
AIPS	MetG, LuxS, GcvP, Hpt, PMI3524	MtnN, CysK, LuxS, MetB, MnmC	MtnN, MnmC, CysK, LuxS, MetB	MtnN, CysK, LuxS, MetB, MnmC
AIPL	LuxS, ThrA MetH, MetL, PMI0028	MnmC, MtnN, LuxS, PMI3678, TrmA	MnmC, LuxS, MtnN, ThrA MetH	PMI3678, ThrA, MetH, PMI0028, PMI0626
QSPO	PMI1345, PMI1344, TrpE, PabA, PabB	YajC, PMI1345, GadC, RibD, PMI2708	PMI1345, PMI1344, TrpE, PabA, PabB	KdpE, Hfq, FlhD, FlhC, PMI1345
QSPH	PMI1345, GadC, TrpE, FlhD, FlhC	KdpE, Ffh, KdpD, LepB, FtsY	OppA, MppA, OppA2, OppD, OppC	FlhD, FlhC, PMI1423, AroF, AroG
QSPV	FliF,FliK,FlgG, FlgC, FlgI	RpoS, Eno, Irp, Pgm, PMI3678	CheY, PykA, PykF, Tal, FliN	RpoS, Eno, PMI3678, FliC, CsrA
GPIN	PolA, GuaA, DnaK, MetG, RecA	PolA, PMI3678, PMI1033 PMI2007, RpoS	PolA, PMI3678, RcsC, DnaK, GuaA	PMI2375 PMI2723, PMI0739, PMI3495, PMI2629

In most of the cases, at least three or two of the centrality measures unanimously brings out the same protein. These proteins are the ones reflected to be important through each SPIN analysis. For instance, AIPS has MetG and MtnN as the top rankers while LuxS and MnmC turns out to be important for AIPL. Others like QSPO, QSPH and QSPV have YajC, PMI1345, Hfq, RpoS, flagellar proteins of the flh and fli operon and some other two-component systems proteins like CheY and KdpE as important rankers. The functions of these proteins are mentioned in Table 2. The top ranking proteins for each of these five SPINs have been reflected in Fig. 1 with Venn diagrams along with the common topmost rankers across all the five SPINs. It is worthwhile to note that the QSPO had a tightly packed network being formed with every protein almost connected to every other, a phenomenon related as high clustering coefficient. With respect to the above analyses of the individual interactomes of the SPIN, an idea about the importance of these proteins in their individual SPIN and finally across all SPIN could be obtained. However, for a drug to be effective globally, the indispensability issue of these proteins needs to be taken care of to tackle the MDR P. mirabilis. Thus, a broader picture with respect to the whole genome proteins of P. mirabilis was then examined to address the concern.

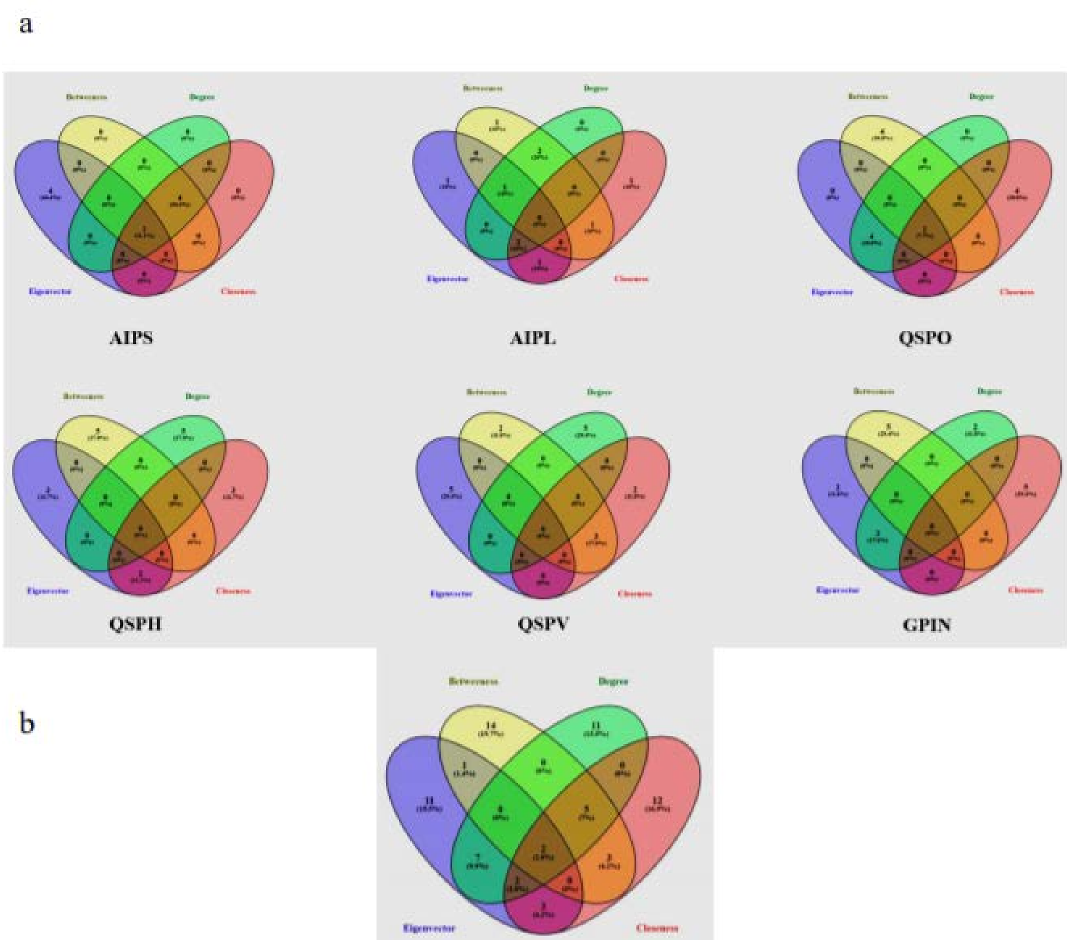


Figure 1: Venn diagram representation for P. mirabilis. a] The five top rankers of DC, CC, BC and EC parametric analyses of five individual SPIN and GPIN and b] The common centrality measures of all interactomes.

The complete GPIN

In an attempt to analyse the type of network being constructed from the empirical and theoretical results of physical and functional interactions amongst proteins laid down in STRING, we have observed the degree distribution of GPIN to be exponential showing a non-linear preferential attachment nature (Fig. 2a) [34]. Hereafter, in order to get an idea of the influential ones from the barrage of proteins involved in the five individual SPIN, we have performed a k-core analysis for them (Fig. 2b). Notably, the innermost core was 154th shell and had genes like thrA, cysK, metG, metL, trpE, rpoS, eno, etc. which have already been reflected from the four network centrality analyses of the SPINs (Table 1). Additionally, it is to be noted that top 5 EC and DC measures of the GPIN also had their position in the innermost 154th core, thereby indicating their importance in the global scenario. Other important genes e.g. luxS, PMI1345 from the k-core analyses were found in the 139th shell. The latter category was found to have direct involvement in QS. For the purpose of classification of the proteins based on their functional role and region in the network space of P. mirabilis, we have performed a cartographic analysis for the GPIN. Noticeably, the R6 quadrant had the top 5 proteins belonging to either the innermost 154th core or almost close to the 139th core containing most of the proteins related to QS. These are GltB and PMI3678 for the former and PMI3348, PMI0587 and PMI3517 for the latter.

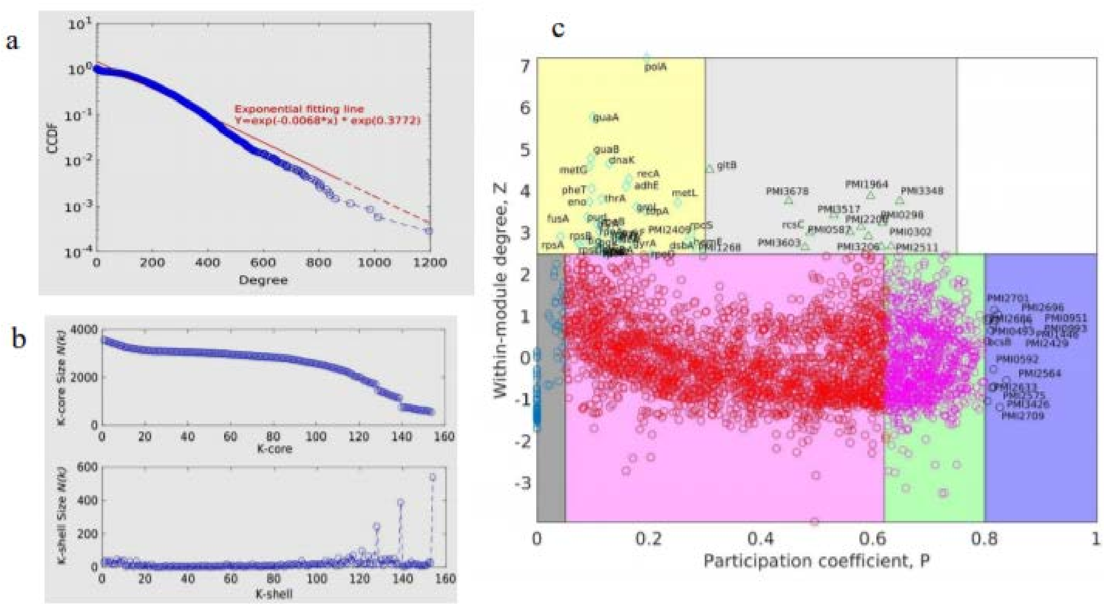


Figure 2: a] The degree distribution of the proteins from the GPIN of P. mirabilis b] Distribution of the kcore (top) and k-shell (bottom) sizes for the set of proteins from the GPIN of P. mirabilis. c] Cartographic representation for classification of proteins from the GPIN of P. mirabilis based on its role and region in network space

Moreover, upon looking deep into EC classification of R6 quadrants, all top 5 proteins, namely PolA, GuaA, DnaK, MetG and RecA were from the innermost 154th core. Furthermore, analysis after sorting of module followed by R quadrant, k-core followed by either module or EC measures, all revealed the proteins to be mostly belonging to the R6 or R5 categories, besides their 154th or 139th core classification (Data not shown). It is worthwhile to mention here that a similar sorting analyses of BC with respect to Quadrant and k-core had revealed proteins mostly from R2 or R3, none of them occupying the innermost 154th core, except RplP and RpoS. **Table 2: Functions of the topmost proteins of individual P. mirabilis networks with respect to network centrality measures.**

Protein name	Description of function
MetG	Is required not only for elongation of protein synthesis but also for the initiation of all mRNA translation through initiator tRNA(fMet) aminoacylation.
MtnN	Catalyzes the irreversible cleavage of the glycosidic bond in both 5'- methylthioadenosine (MTA) and S-adenosylhomocysteine (SAH/AdoHcy) to adenine and the corresponding thioribose, 5'-methylthioribose and Sribosylhomocysteine, respectively.
LuxS	Involved in the synthesis of autoinducer 2 (AI-2) which is secreted by bacteria and is used to communicate both the cell density and the metabolic potential of the environment. The regulation of gene expression in response to changes in cell density is called quorum sensing. Catalyzes the transformation of S-ribosylhomocysteine (RHC) to homocysteine (HC) and 4,5-dihydroxy-2,3-pentadione (DPD).
MnmC	Catalyzes the last two steps in the biosynthesis of 5-methylaminomethyl-2- thiouridine (mnm5s2U) at the wobble position (U34) in tRNA. Catalyzes the FAD-dependent demodification of cmnm5s2U34 to nm5s2U34, followed by the transfer of a methyl group from S-adenosyl-L-methionine to nm5s2U34, to form mnm5s2U34.
PMI3678	Catalytic activity
PMI1345	Catalyzes the transfer of the phosphoribosyl group of 5-phosphorylribose-1-pyrophosphate (PRPP) to anthranilate to yield N-(5'-phosphoribosyl)- anthranilate (PRA).
FlhD	Functions in complex with FlhC as a master transcriptional regulator that regulates transcription of several flagellar and non-flagellar operons by binding to their promoter region. Activates expression of class 2 flagellar genes, including fliA, which is a flagellum-specific sigma factor that turns on the class 3 genes. Also regulates genes whose products function in a variety of physiological pathways.
FliF	The M ring may be actively involved in energy transduction.
PolA	In addition to polymerase activity, this DNA polymerase exhibits 5'-3' exonuclease activity.
RplP	Binds 23S rRNA and is also seen to make contacts with the A and possibly P site tRNAs.

Discussion

We have started with the proteins involved in P. mirabilis AI-2 biosynthesis pathway (Data not shown) and derived the AIPS along with AIPL. While the former connects the proteins of the pathway as reported by default in STRING with only 10 interactors, the latter has been formed upon extending those to 50 interactors per protein query. The idea was to incorporate other related proteins having connectivity to the AI-2 whose analysis might give more insight about QS in P. mirabilis. Moreover, it was necessary to have an idea of the robustness of the proteins involved in QS pathways and thus, the tightly connected QSPO was constructed where proteins were found to have a clustering coefficient of almost 1. Again, with the homologous proteins reported to be involved in QS in other species from KEGG database, it was necessary to look into their association with acknowledged QS proteins of P. mirabilis. Thus, QSPH was constructed to take into consideration of this fact and analyse further. Furthermore, with multiple genes and proteins reviewed for the virulence of P. mirabilis [1], including those involved for QS phenomenon, it was necessary to have an interactome QSPV constructed to analyse their interactions and involvement. All these SPIN were constructed to have an understanding of the influential proteins responsible for QS in P. mirabilis. To confirm that the large network, is neither random like that proposed by Erdos and Renyi [39] nor a small-world type proposed by Watts and Strogatz [40], we have observed the connectivity distribution, P(k), of a particular node in the GPIN getting connected to k other nodes, for large values of k. The GPIN roughly followed the power-law and is free of a characteristic scale [41] with an exponential decay of the degree distribution. Initially, we have analysed the constructed GPIN with a k-core topological parameter. Proteins which belong to outer shell have lower k value and thus, reflect limited number of interacting partner proteins. Moreover, proteins which belong to inner k-core/shell are specific ones, highly interacting with each other and thus, can be considered to be the most important ones. Decomposition of this core, decomposes the network and thus, makes this the innermost core. We have found the 154th core as the innermost one having many proteins involved in the biosynthesis of amino acids, including cysteine and methionine, the amino acid precursor of the components of AI-2 biosynthetic pathway. These proteins rank top for most of the EC measures across the other five SPIN as well. Furthermore, the 139th core was on focus due to its nearby proximity to the innermost core and comprising most of the proteins directly involved in QS. Such inner core member proteins are highly robust, central and thus highly interactive in nature [42]. Our analyses till this far revealed LuxS and PMI1345 to be the prominent EC proteins in the 139th core of the genome. Interestingly, only PolA and RplP, top rankers of BC measures, made it to the innermost 154th core compared to the other topmost EC proteins in that core. This probably reflects the importance of EC measure to reveal the prominent stakeholders of the machinery responsible for the very survival and probably virulence of the organism. Any effective drug target should, thus be selected from this core group with high EC rank. A further delving deep into the functional connectivity of the modules formed in network topological space reinforced our findings this far. The non-hubs and the hubs formed the functional groups of R1-4 and R5-7, respectively. With high P values and zscore, the kinless hubs nodes (R7) are supposed to be important in terms of functionality, which has high connection within module (P) as well as between modules (z). With the same logic, having the least P and z measures, the ultra-peripheral nodes (R1) occupy the least connecting position in the network followed by the peripheral nodes (R2). These nodes can be pruned easily without much affecting the whole network while decomposing it to reach the core. This is nothing but the outermost shells of the k-core measures (refer previous section). The non-hub connectors (R3) are expected to take part in only a small but fundamental sets of interactions. This is just opposite to those of the provincial hubs class (R5) which have many within-module connections. The non-hub kinless nodes (R4) are those with links homogeneously distributed among all modules. The most conserved in terms of decomposition as well as evolution would be however, those from the connector hubs (R6) with many links to most of the other modules. The system would try to retain these connections as essential ones for their very survival. We have observed mostly R5 and R6 classes of proteins occupying the innermost 154th and the QS-involved 139th cores. Furthermore, the EC measures brings out this importance when compared to other measures of centralities. With the highest number (17) of fimbrial operons reported in any sequenced bacterial species, four P. mirabilis fimbriae, namely, MR/P, UCA, ATF and PMF have shown prominent roles in biofilm formation [43]. The thickness, structure, and the amount of exopolysaccharides produced by some biofilms formed by P. mirabilis are influenced by important acylated homoserine lactones [9]. Moreover, some virulence factors are regulated by QS molecules like acylated homoserine lactones (acyl-HSLs) [44]. Of the two QS types, LuxS is an essential enzyme for AI-2 type which is coded by luxS gene having S-ribosylhomocysteine lyase activity [12]. Acetylated homoserine lactone derivatives modifies the expression of virulence factors of P. mirabilis strains [45]. The flhDC master operon is a key regulator in swarmer cell differentiation in Proteus mirabilis, it is known to cause an increased viscosity and intracellular signals [46]. Furthermore, the extracellular signals can be sensed by two-component regulators such as RcsCRcsB [46]. It is important to note that, we have observed many of these already known genes and proteins, viz LuxS, FlhDC to be reflected from our studies as well. To this end, other genes and proteins, e.g. PMI1345 not reported to have connections with the QS and virulence, have also been unearthed from our study. Thus, it is imperative that an in-depth analysis as that mentioned here would bring out the importance of the proteins unearthed through the process.

Conclusion

This work schematically delineates an approach of figuring out the most influential protein in a system of interacting proteins of P. mirabilis. It deals with the computational framework of building of the theoretical networks comprising the five individual SPIN followed by the network parametric analytical approach of identifying the most interacting protein connected to other important proteins in the concerned phenotypes. This is reinforced by the disintegration of the GPIN to the innermost core of the proteins, essential for virulence and survival followed by analysis of centrality measurements. All these lead to the identification of LuxS and PMI1345 to be the most influential ones amongst a group of other proteins being benefitted through network centrality and decomposition analyses. A further investigation of the GPIN brought forth the proteins of important conserved class, potential enough to be the most important ones and thus, influential amongst the barrage of other proteins of the whole genome of P. mirabilis.

Acknowledgments

The authors wish to acknowledge the support of IMSc, India and Sunway University, Malaysia for the provision of computational facilities. We thank Shweta Singh, Eric Wafula and Frank Roberts for their valuable contribution to an earlier form of the work which metamorphosed to the current state. We also thank Bahaa Abdella for the correct formatting in the latest LaTeX format.

Author contributions

CL conceived the concepts, planned and designed the analyses. Data were generated by SP, MIA and SM and analyzed by CL. Artwork was done by MIA and SP and tabulation was done by SM. CL primarily wrote and edited the manuscript aided by additional help from SP.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Schaffer, J. N., Pearson, M. M. (2015). Proteus mirabilis and urinary tract infections. Microbiology spectrum, 3(5).

[2] Jacobsen, S. M., Shirtliff, M. E. (2011). Proteus mirabilis biofilms and catheter-associated urinary tract infections. Virulence, 2(5), 460-465.

[3] https://emedicine.medscape.com/article/226434-overview#a6

[4] Wang, J. T., Chen, P. C., Chang, S. C., Shiau, Y. R., Wang, H. Y., Lai, J. F., ... Lauderdale, T. L. Y. (2014). Antimicrobial susceptibilities of Proteus mirabilis: a longitudinal nationwide study from the Taiwan surveillance of antimicrobial resistance (TSAR) program. BMC infectious diseases, 14(1), 486.

[5] Luzzaro, F., Brigante, G., DAndrea, M. M., Pini, B., Giani, T., Mantengoli, E., ... Toniolo, A. (2009). Spread of multidrug-resistant Proteus mirabilis isolates producing an AmpC-

type -lactamase: epidemiology and clinical management. International journal of antimicrobial agents, 33(4), 328-333.

[6] Eliopoulos GM, Maragakis LL, Perl TM. Acinetobacter baumannii: epidemiology, antimicrobial resistance, and treatment options. Clinical Infectious Diseases. 2008; 46(8):1254-63.

[7] Czerwonka, G., Guzy, A., Kaua, K., Grosicka, M., Daczuk, M., Lechowicz, ., ... Kaca, W. (2016). The role of Proteus mirabilis cell wall features in biofilm formation. Archives of microbiology, 198(9), 877- 884.

[8] Rutherford, S. T., Bassler, B. L. (2012). Bacterial quorum sensing: its role in virulence and possibilities for its control. Cold Spring Harbor perspectives in medicine, 2(11), a012427.

[9] Stankowska, D., Czerwonka, G., Rozalska, S., Grosicka, M., Dziadek, J., Kaca, W. (2012). Influence of quorum sensing signal molecules on biofilm formation in Proteus mirabilis O18. Folia microbiologica, 57(1), 53-60.

[10] Holden, M. T., Ram Chhabra, S., De Nys, R., Stead, P., Bainton, N. J., Hill, P. J., ... Rice, S. (1999). Quorum sensing crosstalk: isolation and chemical characterization of cyclic dipeptides from Pseudomonas aeruginosa and other gram negative bacteria. Molecular microbiology, 33(6), 1254-1266.

[11] Campbell, J., Lin, Q., Geske, G. D., Blackwell, H. E. (2009). New and unexpected insights into the modulation of LuxR-type quorum sensing by cyclic dipeptides. ACS chemical biology, 4(12), 1051-1059.

[12] Schneider, R., Lockatell, C. V., Johnson, D., Belas, R. (2002). Detection and mutation of a luxSencoded autoinducer in Proteus mirabilis. Microbiology, 148(3), 773-782.

[13] Rauprich, O., Matsushita, M., Weijer, C. J., Siegert, F., Esipov, S. E., Shapiro, J. A. (1996). Periodic phenomena in Proteus mirabilis swarm colony development. Journal of bacteriology, 178(22), 6525-6538.

[14] Pan, A., Lahiri, C., Rajendiran, A., Shanmugham, B. (2015). Computational analysis of protein interaction networks for infectious diseases. Briefings in bioinformatics, 17(3), 517-526.

[15] Lahiri, C., Pawar, S., Sabarinathan, R., Ashraf, M. I., Chand, Y., Chakravortty, D. (2014). Interactome analyses of Salmonella pathogenicity islands reveal SicA influential for virulence. Journal of theoretical biology, 363, 188-197.

[16] Jeong, H., Mason, S. P., Barabsi, A. L., Oltvai, Z. N. (2001). Lethality and centrality in protein networks. Nature, 411(6833), 41.

[17] Seidman, S. B. (1983). Network structure and minimum degree. Social networks, 5(3), 269-287.

[18] Guimera, R., Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. nature, 433(7028), 895.

[19] Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research, 45(D1), D353-D361

[20] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M. (2015). KEGG

as a reference resource for gene and protein annotation. Nucleic acids research, 44(D1), D457-D462.

[21] Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., ... Ong, Q. (2015). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic acids research, 44(D1), D471-D480.

[22] Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., ... Jensen, L. J. (2016). The STRING database in 2017: quality-controlled proteinprotein association networks, made broadly accessible. Nucleic acids research, gkw937.

[23] Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., ... Hanspers, K. (2007). Integration of biological networks and gene expression data using Cytoscape. Nature protocols, 2(10), 2366.

[24] Bastian, M., Heymann, S., Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. Icwsm, 8, 361-362.

[25] Diestel R. Graph theory. Springer-Verlag Berlin and Heidelberg, 2000.

[26] Koschtzki, D., Schreiber, F. (2004, October). Comparison of Centralities for Biological Networks. In German Conference on Bioinformatics (pp. 199-206).

[27] zgr, A., Vu, T., Erkan, G., Radev, D. R. (2008). Identifying gene-disease associations using centrality on a literature mined gene-interaction network. Bioinformatics, 24(13), i277-i285.

[28] Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., ... Bagos, P. G. (2011). Using graph theory to analyze biological networks. BioData mining, 4(1), 10.

[29] Tang, Y., Li, M., Wang, J., Pan, Y., Wu, F. X. (2015). CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. Biosystems, 127, 67-72.

[30] Oliveros, J. C. An interactive tool for comparing lists with Venns diagrams (20072015).

[31] MATLAB and Statistics Toolbox Release 2010b, The MathWorks, Inc., Natick, Massachusetts, United States.

[32] Guimera, R., Amaral, L. A. N. (2005). Cartography of complex networks: modules and universal roles. Journal of Statistical Mechanics: Theory and Experiment, 2005(02), P02001.

[33] Rosvall, M., Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. PloS one, 6(4), e18209.

[34] Vzquez, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Physical Review E, 67(5), 056104.

[35] Claret, L., Hughes, C. (2000). Rapid Turnover of FlhD and FlhC, the Flagellar Regulon Transcriptional Activator Proteins, during Proteus Swarming. Journal of bacteriology, 182(3), 833-836.

[36] Wang, M. C., Chien, H. F., Tsai, Y. L., Liu, M. C., Liaw, S. J. (2014). The RNA chaperone Hfq is involved in stress tolerance and virulence in uropathogenic Proteus mirabilis. PLoS One, 9(1), e85626. https://doi.org/10.1371/journal.pone.0085626.

[37] Rhoads, D. B., Laimins, L., Epstein, W. O. L. F. G. A. N. G. (1978). Functional organization of the kdp genes of Escherichia coli K-12. Journal of bacteriology, 135(2), 445-

452.

[38] Pearson, M. M., Sebaihia, M., Churcher, C., Quail, M. A., Seshasayee, A. S., Luscombe, N. M., ... Hauser, H. (2008). Complete genome sequence of uropathogenic Proteus mirabilis, a master of both adherence and motility. Journal of bacteriology, 190(11), 4027-4037.

[39] Erds, P., Rnyi, A. (1960). On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci, 5, 17-61.

[40] Watts, D. J., Strogatz, S. H. (1998). Collective dynamics of small-world networks. Nature, 393(6684), 440.

[41] Albert, R., Jeong, H., Barabsi, A. L. (2000). Error and attack tolerance of complex networks. Nature, 406(6794), 378.

[42] Alvarez-Hamelin. I, Dall'Asta. L, Barrat. A. and Vespignani. A, Advances in Neural Information Processing Systems 18, MIT Press, Cambridge, MA, 2006; 41-50.

[43] Scavone, P., Iribarnegaray, V., Caetano, A. L., Schlapp, G., Hrtel, S., Zunino, P. (2016). Fimbriae have distinguishable roles in Proteus mirabilis biofilm formation. FEMS Pathogens and Disease, 74(5), ftw033.

[44] Henke, J. M., Bassler, B. L. (2004). Bacterial social engagements. Trends in cell biology, 14(11), 648-656.

[45] Stankowska, D., Kwinkowski, M., Kaca, W. (2008). Quantification of Proteus mirabilis virulence factors and modulation by acylated homoserine lactones. J Microbiol Immunol Infect, 41(3), 243-253.

[46] Fraser, G. M., Hughes, C. (1999). Swarming motility. Current opinion in microbiology, 2(6), 630- 635.

Supervised machine learning algorithms for accurately classifying cancer types from complex gene expression systems

Shrikant Pawar[1*]

[1]Department of Computer Science, Georgia State University, 34 Peachtree Street, 30303, Atlanta, GA, USA.

[*]Correspondence should be addressed; E-mail: spawar2@gsu.edu

Abstract:

Different supervised machine learning algorithms have been implied in classifying cancer types and complex gene expression patterns amongst these types. Microarray technique may not be an exhaustive but definitely one of the accurate techniques of gene expression analysis. Application of supervised machine learning algorithms on different normalization algorithms like data driven harr fisz, Z score and principal component analysis can enhance the analysis sensitivity. This paper focuses on enhancement pipeline generated on a lung cancer dataset from gene expression omnibus analyzed with selected algorithms for classifying cancer types and gene expression patterns.

Introduction:

Hybridization of RNA from tissues to the arrays is an effective technique of analyzing gene expression patterns [1]. Different array spots provide different fluorescence intensity values for each gene which can be analyzed with different Bioconductor package in R to perform normalization and statistical analysis. Supervised machine learning techniques like Support Vector Machine (SVM) categorizes new samples based on supervised learning of training data. A hyperplane is defined which forms a minimum distance in training sample which is used for classification [2]. SVM makes classifications based on non-probabilistic binary linear classifier [3]. It has been shown that SVMs can significantly increase accuracy compared to traditional query refinement schemes [3]. This paper summarizes some supervised machine learning techniques, which can input a one channel microarray data and classify cancer types and complex gene expression systems. The pipeline for such analysis as depicted in Figure 1.

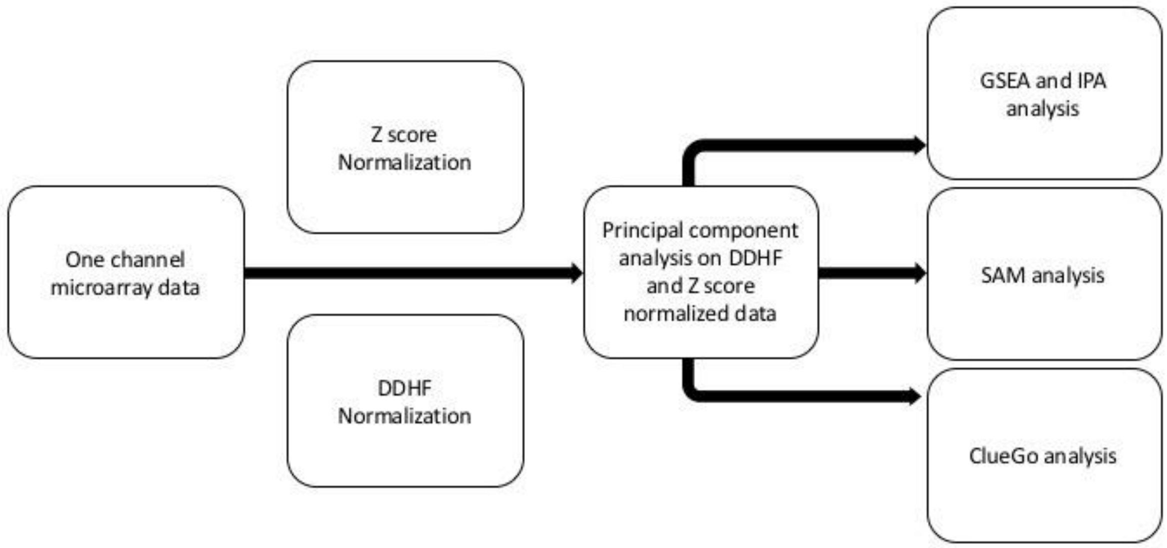


Figure 1: Pipeline for the analysis.

Materials and Methods:

a. Data Collection:

One channel microarray lung cancer samples (10,000 genes per channel with series id GPL590) were downloaded from Gene Expression Omnibus (GEO) database. Similar number of normal lung tissue microarray data accompanied cancer data.

b. Normalization:

Normalization is needed to decrease the standard deviation and variance of the data which was achieved with two methods, data driven harr fisz (DDHF) transformation and the Z score transformation. DDHFm achieves very good variance stabilization of microarray data with replicates and produces transformed intensities that are approximately normally distributed [4]. Z ratios provide a relative measure of significant gene expression changes in pair wise group comparison. The advantages of Z ratios are that they are directly comparable among different experiments, rapidly calculated, and show good agreement with more complex statistical analysis like SAM analysis [5]. The DDHF package was downloaded from Bio-conductor and was used in R environment to run DDHF transformation (DDHFm(data)) [4]. The distributions of raw intensity plots for each of the treatments were generated in Mathematica and are shown in Figure 2 (a-d).

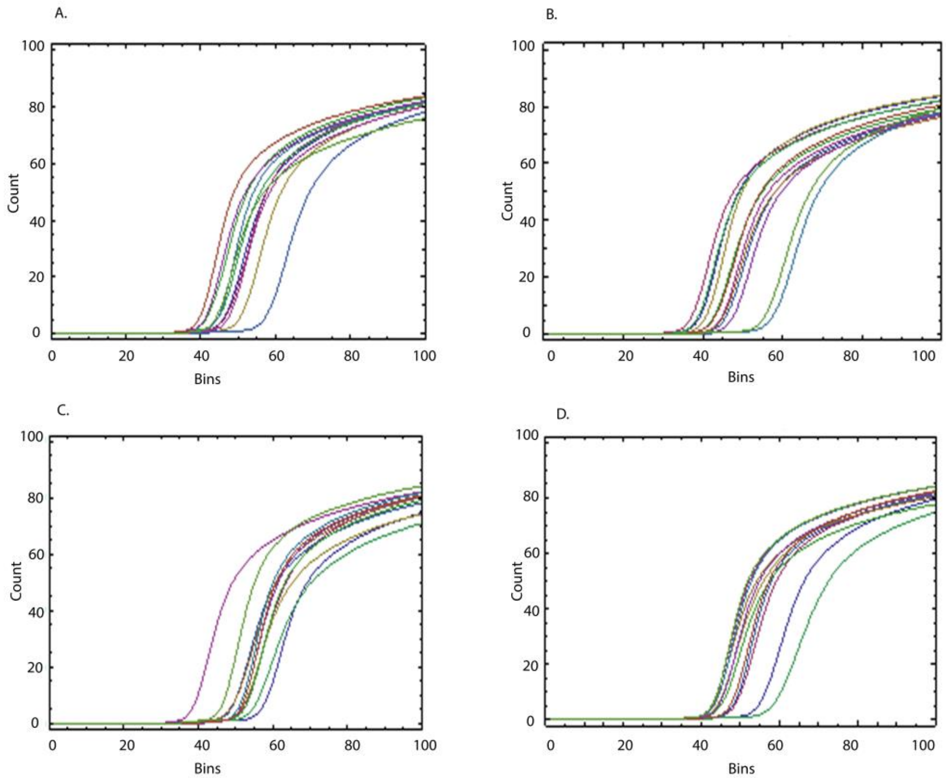


Figure 2: distributions of raw intensity plots for each of the treatments.

The distribution of normalized intensity data is shown in Figure 3 (a-d).

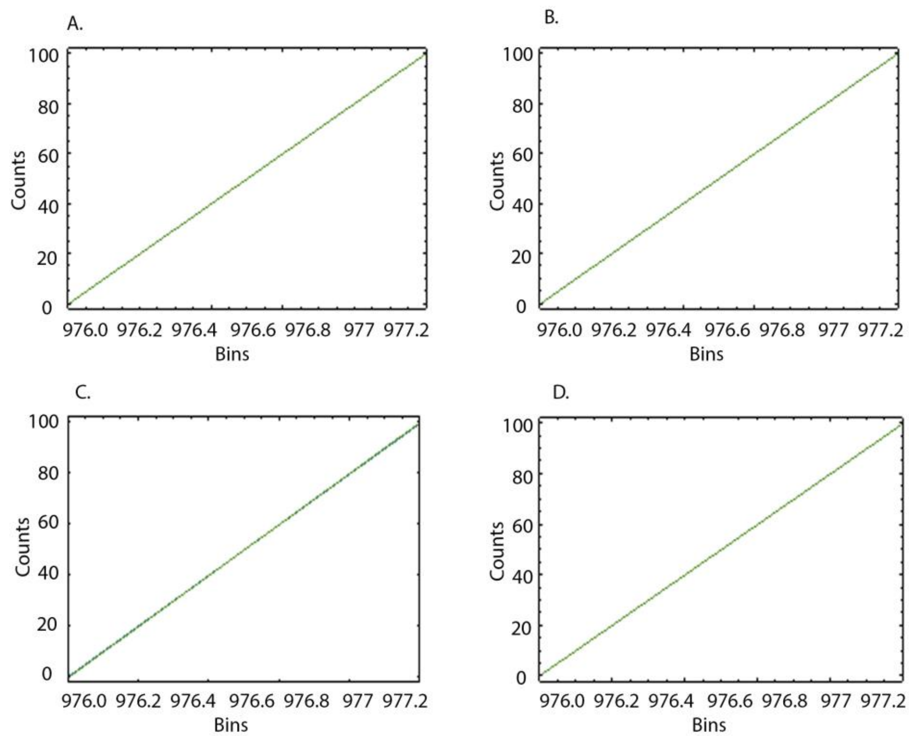


Figure 3: distribution of normalized intensity data.

Chips before normalization had ample inadequate variance, which was stabilized after DDHF normalization. The probability associated with a student's paired t-test (two tailed distribution) was taken and the genes with p values less than or equal to 0.05 were considered for further processing [9, 10, 11]. DDHF transformation estimates the mean variance relation as a part of the process of stabilization and brings the distribution closer to normality. If we have sequence of positive random variables xi with mean μi and variance σ2i with some monotone relation between the mean and variance as σ2i = h(μi), then the DDHF transformation will work best when the underlying μi forms a piecewise constant sequence [4]. Simulations have shown that DDHFm performs very good variance stabilization and also produces intensities that have distribution much closer to the Gaussian transformations [4].

Z score normalization: The Z score transformation was applied on the raw data to determine the effectiveness of DDHF normalization [5]. Average and standard deviation of logarithm to base 10 for the raw intensities was calculated [12]. Each of the intensity value was then subtracted from the average, and the result was divided by the standard deviation to find z scores for each gene. To calculate Z ratios, the average for each gene was taken on Z scores across each of the first and next chips. The first average was subtracted from the second. Z ratio was then calculated by dividing each of the difference by standard deviation of difference. The Z ratios with values more than +1.96 were considered up regulated and values more than -1.96 were considered down regulated. Fold change calculates significant changes in gene expression derived from globally normalized data. It is performed by calculating the ratio of average of all the measurements from one condition over another [6]. In our analysis, we have applied a stringent threshold where the fold changes with values more than 2 were considered up regulated and less than 0.5 were considered down regulated [13, 14]. Z score transformation was also found to be effective compared to DDHF normalization with all the chips in each of the treatments falling in a single linear pattern as seen in Figure 3 (a-d).

Principal component analysis (PCA): PCA was performed with CLC main work bench 6 software [7]. PCA is a method that reduces data dimensionality by performing a covariance analysis between factors. It is suitable for data sets in multiple dimensions such as a large experiment in gene expression. To characterize the trends exhibited by data, PCA extracts directions where the cloud is more extended [7]. It forms an orthogonal space to the first component, reducing the multidimensional cloud into a two-dimensional space [8]. PCA on genes provide a way to identify predominant gene expression patterns. When applied on conditions, PCA will explore correlations between samples or conditions [8]. The PCA plots on Z score and DDHF transformed data were generated and are shown in Figure 4 (panel A).

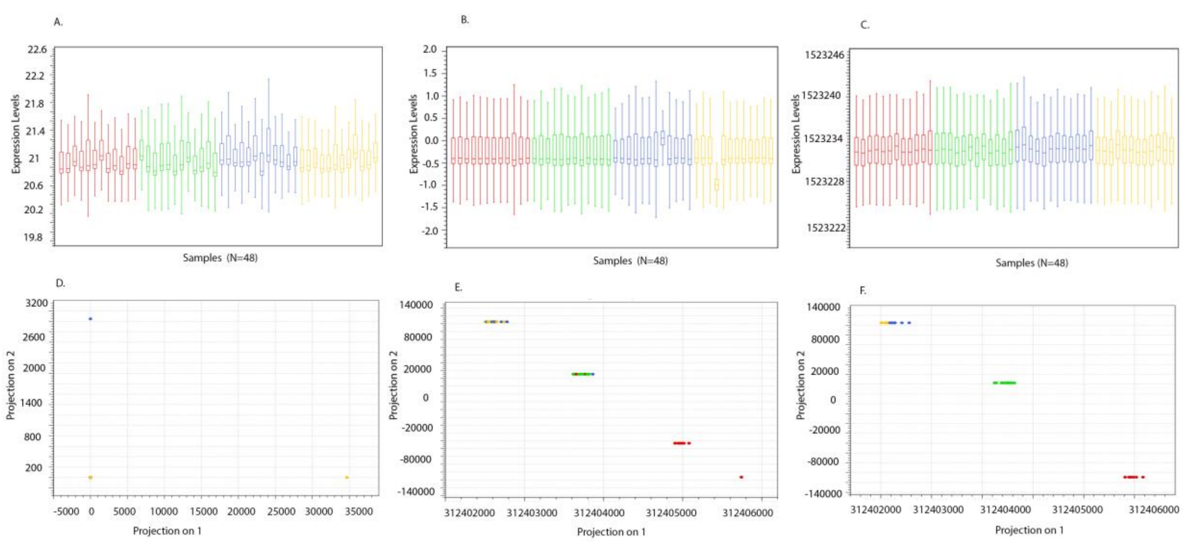


Figure 4: Panel A: Z score and DDHF transformed data, Panel B: PCA plots.

PCA should identify groups of all the replicates in one component. It should group all the replicates from each of the treatments in one. Although, Z score transformation within treatments was very stable, the between treatment, Z score transformation seems little inefficient with 2 replicates falling as outliers as seen in Figure 4 (panel B). Moreover, PCA was not able to cluster replicates in similar clusters with Z score values. Within DDHF, we found that with PCA on DDHF normalized data, some replicates overlapped between treatment categories (PCA on DDHF overlap plot), therefore some of the replicates were dropped as indicated below in order to develop a non-overlapping experimental design (DDHF non-overlap plot).

c. Machine Learning Analysis:

SVM is widely used as a supervised learning classifier in the classification area [15]. SVM analysis was performed in a linear kernel mode by implementing Python sklearn toolbox [16]. We found a high accuracy, sensitivity, 95 percent confidence, positive and negative prediction values and specificity for all the tested datasets. A sample lung cancer GSM159355 used as a query sample provided a prediction of 94.88 percent to be a lung cancer type. For different series, accuracies of 93.08, 92.11, 92.95, 91.05, 92, 93.10, 94.99 and 90.11 percent were produced. Randomly chosen (k-1)/k chips (some are cancer, while others are normal) were selected for training the classifier and the remaining 1/k sequences were used for testing. A three, five and tenfold validation was performed for SVM analysis.

d. Setting up a server environment for decreasing computation time for log complexity algorithms:

Although, the run-time of PCA, ddhf and Z scores are exhaustive and slow with log complexities, we didn't have any run-time issues with our data on an IBM system x3850 X5 server. A Linux server hosts the application and database. A MySQL as the relational database management system and Python as the scripting language has been utilized. The hardware requirements of suite on the server side are moderate. The server we utilized is CentOS 6.7 64-bit, 6x IBM System x3850x5, Intel Xeon Processor E7-4850, 4 CPUs (10 cores per CPU), 2.0 GHz processors, 512 GBRAM and 2 TB of scratch storage for jobs.

Results and Conclusions:

We have tried to follow a general pipeline of analyzing the microarray data with different supervised machine learning algorithms. The expansion of training set is needed to improve prediction accuracies. Also increasing cancer types (breast, brain, liver etc.) can make this pipeline usable for different groups. Other supervised algorithms like divisive clustering which is a ’top down’ approach, where all the observations start in one cluster and splitting occurs recursively as one moves down the hierarchy should be implied against SVM [17]. Divisive cluster analysis can be performed using function ’diana{cluster}’ from cluster package in R [18]. This algorithm constructs a hierarchy of clustering’s, starting with one large cluster containing all ’n’ observations with divisions until each cluster contains only a single observation [18]. K means method aims to partition the points into K groups, such that the sum of squares from points to the assigned cluster centre’s is minimized, it at-most tries to put all cluster centre’s at the mean of their Voronoi sets [19]. For a set of observations {x1, x2, ..., xn}, K means clustering aims to partition the ’n’ observations into K sets, S ={S1, S2, ..., Sk}, so as to minimize the within-cluster sum of squares (WCSS), where μi is the mean of points in set Si [20]. It can be summarized by following equation:

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i$$

Hierarchical agglomerative clustering is a ’bottom up’ approach, where each observation starts in its own cluster, followed by a continuous merging with upwards hierarchical movement [17]. The choice of an appropriate metric is important in influencing the shape of the clusters. Distance plays an important factor in cluster distribution and various distance techniques such as euclidean, squared euclidean, manhattan, maximum and mahalanobis distances have been implied in such clustering [17]. Euclidean distance between two vectors ’a’ and ’b’, each with ’i’ points can be computed as follows [17]:

$$\|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2} \|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2}$$

Funding:

No external funding for developing this suite has been utilized.

Competing interests:

The authors declare that they have no competing interests.

Author's contributions:

SP conceived, designed the study and critically revised the manuscript. SP developed, tested the software and also did the setup of SQL database and server space.

Acknowledgements:

Not applicable.

References:

1. Quackenbush J. (2002). Microarray data normalization and transformation. Nature Genetics, 496 - 501. doi:10.1038/ng1032.

2. OpenCV. (2014). Introduction to Support Vector Machine. Retrieved from http:/

223http://docs.opencv.org/2.4/doc/tutorials/ml/introductiontosvm/introductiontosvm.html.

3. Ben-Hur, Asa, Horn, David, Siegelmann, Hava, and Vapnik, Vladimir (2001). "Support vector clustering". Journal of Machine Learning Research, 2: 125–137.

4. E S Motakis, P.F.G.A.R. G P Nason: Variance stabilization and normalization for one-color microarray data using a data-driven multiscale approach. Bio Informatics 22 (2006)

5. Chris Cheadle, W.J.F.K.G.B. Marquis P Vawter: Analysis of microarray data using z score transformation. J Mol Diagn, 73–81 (2003)

6. Yang YH, L.P.L.D.P.V.N.J.S.T. Dudoit S: Normalization for cdna microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res (2002)

7. Battke F, N.K.: Mayday seasight: Combined analysis of deep sequencing and microarray data. PLoS ONE (2011

8. Raychaudhuri S, A.R.B. Stuart J M: Principal components analysis to summarize microarray experiments: application to sporulation time series. Pacific Symposium on Biocomputing (2000)

9. Pawar S, Davis C, Rinehart C (2011) Statistical analysis of microarray gene expression data from a mouse model of toxoplasmosis. BMC Bioinform 12(Suppl 7):A19

10. Lahiri C, Pawar S, Sabarinathan R, Ashraf MI, Chand Y, Chakravortty D (2014) Interactome analyses of Salmonella pathogenicity islands reveal SicA indispensable for virulence. J Theor Biol 363:188–197

11. Ashraf M, Ong S-K, Mujawar S, Pawar S, More P, Paul S, Lahiri C (2018) A side-effect free method for identifying cancer drug targets. Sci Rep. https://doi.org/10.1038/s41598-018-25042-2

12. Pawar S, Ashraf M, Mujawar S, Mishra R, Lahiri C (2018) In silico Identification of the indispensable quorum sensing proteins of multidrug resistant Proteus mirabilis. Front Cell Infect Microbiol 8:269

13. Pawar S, Ashraf M, Mehata K, Lahiri C (2017) Computational identification of indispensable virulence proteins of Salmonella Typhi CT18. Curr Top Salmonella Salmonellosis. https://doi.org/10.5772/66489

14. Lahiri C, Shrikant P, Sabarinathan R, Ashraf M, Chakravortty D (2012) Identifying indispensable proteins of the type III secretion systems of Salmonella enterica serovar Typhimurium strain LT2. BMC Bioinform 13(Suppl 12):A10

15. Cortes C, V.V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995)

16. Fabian Pedregosa, Ga¨el Varoquaux: Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12, 2825–2830

17. Rokach Lior, O.M.: Clustering methods. Data mining and knowledge discovery handbook, 321–352 (2005

18. Martin Maechler, A.S.M.H.K.H. Peter Rousseeuw: cluster: Cluster analysis basics and extensions. R package version 2.0.6 (2017)

19. Hartigan J A, W.M.A.: A k-means clustering algorithm. Applied Statistics 28, 100–108 (1979)

20. Kriegel, S.E.Z.A. Hans Peter: The (black) art of runtime evaluation: Are we comparing algorithms or implementations? Knowledge and Information Systems 52, 341–378 (2016)

A Comparative Study of Various AI Based Breast Cancer Detection Techniques

Sujatha Alla, Leili Soltanisehat, and Aysen Taylor
Old Dominion University, Norfolk VA 23529, USA
Salla001@odu.edu, Lsolt001@odu.edu, **akahy001@odu.edu**

Abstract
As Artificial Intelligence (AI) and learning techniques are making great leaps in the detection of abnormalities in tumors, it is important to see which technique is most efficient in terms of accuracy. This, in turn, will allow for timely detection, early treatment, and reduce physician burnout. Supervised learning, deep learning, and data visualization techniques are some of the common methods used to improve the accuracy of cancerous tumor detection. Choosing the best detection method can be a complex problem due to different characteristics of the methods and the features of the data. This paper aims to analyze the literature of breast cancer detection and develop a comprehensive comparison between various models that can differentiate between malignant and benign breast tumors. The results will help to determine which method is most accurate for diagnosis of the cancer stage. Furthermore, studying the effectiveness of implementation of AI tools in breast cancer detection helps to determine the degree to which the future behavior of the system can be anticipated in terms of Predictability, Reliability (Robustness), Dependability, and trust (future behavior of the system) in the automated detection system. We aim to study the robustness of AI-based machine learning algorithms that can read into the breast cancer data, analyze the features, and predict the future outcomes whether the tumor is going to be cancerous or not. In order to have validated results, the real datasets will be extracted from The Cancer Imaging Archive (TCIA) and tested with the chosen methods.

Keywords
Artificial Intelligence, Breast Cancer Detection Models, Machine Learning, Deep Learning

1. Introduction
Artificial intelligence (AI) is one of the current megatrends emerging from the broader digitization of society and the economy. AI-supported breast cancer screening is gaining popularity. Initial use cases have been found for AI-supported systems that enhance care – for instance, in the development of various risk models to assess the elevated risk for patients suffering from chronic breast cancer while adhering to the clinical guidelines recommended by American Cancer Society (ACS) (Wright et al., 2011). Screening breast cancer is a complex process and choosing appropriate machine learning technique would be another aspect of complexity involved in this whole automation process. However, the management and resolution of such an immensely complex system is difficult and may require a system of systems approach. In addition, involving large datasets may offer a resolution in time, but the scale of the task is daunting (Jupp 2018). Cognitive systems should be able to identify women at elevated risk and offer more intensive screening if warranted (Tilanus-Linthorst, Bartels, Obdeijn, & Oudkerk, 2000). Various machine-learning models can help care providers to efficiently screen cases, evaluate them with greater precision, and make informed decisions. Reliably identifying and detecting these abnormalities would save all stakeholders – patients and providers alike – a great deal of time, money, and effort. Such a system can systematically identify and correct errors while avoiding unnecessary or ineffective interventions. In addition to machine-learning technologies – which can likewise track developments, recognize patterns, and classify them – artificial intelligence is able to apply what it learns, to new situations in order to perform a more accurate prediction (Purdy & Daugherty).

1.1. Benefit of this Study

Cancer incidence continues to be a major health problem possibly because cancer is a complex system comprising many agents that interact in a non-linear manner resulting in many possible outcomes (Jupp 2018). Breast cancer is a major terminal disease that occurs largely among females and it is known as the second most diagnosed cancer and the second leading cause of cancer death in women (Siegel, Miller, & Jemal, 2018). The American Cancer Society (ACS) predicted that 246,600 new cases of breast cancer would be diagnosed in women with 40,890 deaths during 2016. Early detection, in conjunction with appropriate screening, can decrease mortality associated with breast cancer (Saslow et al., 2007). AI-based machine learning models promise the early detection by predicting the fatal tumors. The increasing availability of healthcare data and rapid development of big data analytic methods have made possible the recent successful applications of AI in breast cancer detection (Himes, Root, Gammon, & Luthy, 2016). Guided by relevant clinical questions big data analytic methods, and powerful AI techniques can unlock clinically relevant information hidden in the massive amount of data, which in turn can assist clinical decision-making (Dilsizian & Siegel, 2014). Nevertheless, it is important to consider the fact that breast cancer (specially most common form; sporadic breast Cancer) is a complex disease due to the due to multiple unknown etiologies and the existence of interdependent agents/factors, which cause the cancer (Ritchie et al. 2001). The predictions are made based on complex relationships between these factors which makes interpretability of the results even more difficult (Jerez et al. 2010). Because of the inherent complexity of cancer formation, detection and the cure process, there is much interest in using machine learning techniques to help deal with this complexity. Furthermore, considering different new detection methods, choosing appropriate machine-learning techniques, which boost the accurate cancer detection and help to bias correction in the prediction phase, would be another aspect of the complexity.

2. History and Background

The Center for Managing Chronic Diseases (CMCD) in the United States of America defined a chronic disease as a long-standing condition that can be controlled but not cured (McGuire, 2016). Normally, radiologists use risk-calculating models to determine the perceived risk of a breast cancer patient by using their clinical data (Saslow et al., 2007). However, using risk-calculating models is time intensive. To use these models, healthcare providers must collect extensive family history, often necessitating more than 1 office visit to allow patients time to collect information from extended family. It would not be feasible to calculate lifetime risk for every patient. Therefore, healthcare providers opt to rely on software-based models to identify who needs more intensive risk assessment (Saslow et al., 2007; Tilanus-Linthorst et al., 2000).

Most of enhancements in computer modeling, particularly in the image analysis field in health care, have focused on feature engineering, essentially asking a computer to evaluate explicit features specified by experts. This permits the algorithms to detect abnormalities or predict specified lesions. In contrast, deep learning is a form of AI that includes machine-learning techniques that perform iterative optimization strategies that are based on a pixel-by-pixel evaluation of the data from images (LeCun, Bengio, & Hinton, 2015). The promise of AI in healthcare is the delivery of improved quality and safety of care and the potential to democratize expertise. Radiology, having converted to digital images more than 25 years ago, is well-positioned to deploy AI for diagnostics. Several studies have shown a considerable opportunity to support radiologists in evaluating a variety of scan types including mammography for breast lesions, computed tomographic scans for pulmonary nodules and infections, and magnetic resonance images for brain tumors including the molecular classification of brain tumors (Dilsizian & Siegel, 2014).

3. Methodology

This section defines the overall approach used to address the research goals. By applying different machine learning techniques to the breast cancer data, this research is trying to answer the following question:

Which learning model proves to be robust and accurate in detecting, diagnosing, and predicting the pathology of a tumor (malignant or benign) by analyzing the breast cancer data?

First, we will look into some studies that already have implemented a learning model to detect the breast cancer and report their results. Secondly, we will take the unstructured breast cancer datasets and apply various machine-learning models to test their efficiency in accurately detecting the breast cancer disease. The result will be an ensured framework used to choose the best and the most accurate method for predicting the cancer disease for the uncertain cases.

3.1. Case study 1

A study by Ehteshami Bejnordi and colleagues report the results of an investigation developed in response to an international contest to have a machine detect sentinel lymph node metastases of breast cancer (Bejnordi et al., 2017).

Two hundred and seventy whole-slide images were provided to the 390 entrants to build their algorithms along with an independent test set of 129 whole-slide images for which the actual diagnosis was obscured. A total number of 23 teams submitted 32 methods for evaluation, nearly 80% using a deep CNN method (Bejnordi et al., 2017). While several teams submitted methods other than deep learning-based algorithms, the deep CNN methods performed significantly better. At least the top 5 algorithms performed as well as pathologists, if not slightly better, with several caveats. Confidence in the algorithms comes from their ability to detect metastases. The top algorithms performed better than the 11 pathologists with time constraint at identifying micro-metastases (tumor cell cluster diameter, 0.2 to <2 mm) (area under the curve for the best algorithm, 0.994; 95% CI, 0.983-0.999 vs mean area under the curve for the 11 pathologists, 0.810; range, 0.738-0.884; $P < .001$)

(Bejnordi et al., 2017). Although micro-metastases are currently being evaluated for their clinical importance, the fact that the algorithms detected these abnormalities at the same rate or better than pathologists is exciting. The algorithms were better than the 11 pathologists with time limits, at detecting noninvasive ductal carcinoma.

3.2. Case study 2

Braunstein (2016), demonstrated the use of machine learning models in decision support systems. Pervasive computing and artificial intelligence put together will result in a stronger framework for the deployment of accessible and approximate medical expert systems. Artificial Intelligent (AI) systems are supposed to support health workers in tasks that rely on data and knowledge manipulation (Braunstein, 2016). Systems resulting from this are usually termed Clinical Decision Support Systems (CDSS) and are useful in alerts and reminders, therapy critiquing and planning, diagnostic assistance, information retrieval, image recognition and interpretation and prescribing decision support system (Lee et al., 2014). Diagnosis Decision Support System (DDSS) requires the patient to enter some required information and then intelligently carry out diagnoses and then respond to a patient with the appropriate set of diagnoses, and then a physician may select the diagnosis relevant to the patient in question (Oyelade, Obiniyi, Junaidu, & Kana, 2017). These decision support systems are driven by reasoning algorithms such as Select and Test (ST). The improved ST was enabled to automatically generate input base on clinical protocols from benchmarking datasets from of Breast Cancer Wisconsin Dataset retrieved from UCI Learning Repository (Oyelade et al., 2017).

In this research, the authors tried to validate the proof of the result obtained by using standard or benchmark dataset for comparing the performance of the implemented medical diagnosis system, which is based on an improved Select, and Test (ST) algorithm. Three datasets were retrieved from their online repositories, and these datasets are namely: Wisconsin Breast Cancer Database (WBCD) authored by Wolberg (1992), Wisconsin Diagnostic Breast Cancer (WDBC), and Wisconsin Prognostic Breast Cancer (WPBC). The WBCS database has 699 instances (as of 15 July 1992), 10 numbers of attributes plus the class attribute. The approach of this research in using these datasets is to generate input from each dataset and then feed the inputs into the medical expert system proposed by the research. This source of input became relevant to this research considering the similarity in the ailment (breast cancer) being diagnosed by the providers of the databases and this research. Hence, based on the given attributes in each dataset, this research retrieved corresponding inputs from the database of acceptable tokens/inputs developed in the course of this research. Since the datasets only lists attributes names and their corresponding weights on each instance, this research embarks on the style of sourcing all symptoms, signs, medical terms modeled in our database which, matches a given attribute, and then assign the weight of the attribute for each instance to be the likelihood of the symptom or sign to be fed as input into the proposed medical expert system.

Their result shows that there was Non-Breast Cancer (benign) diagnosed instances to be 73.40% and Breast Cancer (malignant) diagnosed instances to be 26.60% as against the 458 (65.5%) cases of benign and 241 (34.5%) cases of malignant observed by the authors of the dataset. It was able to compute the specificity of the three databases as 0.89, 0.706, and 0.60 respectively, while that of sensitivity are 0.81, 1.0, and 1.0.

3.3. Case study 3

Another study performed a comprehensive study on the deep-learning-based computer-aided diagnosis (CADx) for the differential diagnosis of benign and malignant nodules/lesions by avoiding the potential errors caused by inaccurate image processing results (e.g., boundary segmentation), as well as the classification bias resulting from a less robust feature set, as involved in most conventional CADx algorithms. To achieve malignancy identification, the conventional design of CADx is often composed of three main steps; feature extraction, feature selection, and classification. These three steps need to be well-addressed separately and then integrated together for the overall CADx performance tuning. The training of an SDAE-based CADx framework can be realized in two steps; the pre-training and supervised training steps.

The authors found that the deep learning architecture holds the advantages of 1) automatic discovery of object features, 2) automatic exploration of feature hierarchy and interaction, 3) relatively simple end-to-end training process, and 4) systematic performance tuning (Cheng et al., 2016).

3.4. Case study 4

Mammography is considered the primary imaging modality for early detection and treatment of breast cancer; however, achieving accurate diagnoses through mammography is often challenging for radiologists due to the difficulty of distinguishing the features of malignant symptoms in various images (Thrall et al., 2018). Consequently, considerable research is being undertaken to develop computer-based applications including various classification models to overcome these challenges. Micro-calcifications are highly correlated with breast cancer (McGuire, 2016), therefore, the aim of this investigation was to evaluate the performance of an innovative deep learning model for classifying breast lesions (Wang et al., 2016). The results demonstrated that deep learning not only enabled accurate segmentation of micro-calcifications but also provided an efficient analysis of their characteristics, leading to a marked improvement in discriminating between benign and malignant breast lesions compared to more standard SVM, KNN, and LDA methods. This may have particular significance for cases in which micro-calcifications are the only indicator of malignant lesions (Wang et al., 2016).

3.5. Data analysis

In this research, two sample raw breast cancer datasets are taken from The Cancer Imaging Archive (TCIA)[1] which is an online accessible source. The dataset is pre-processed, in a way that only features that are relevant to the analysis are considered. We will use python v.3 and Sci-kit package to develop the machine-learning algorithm for three main steps; (1) cleaning the data, (2) developing, fitting, and training the model, and (3) implementing predictive analysis. Since we already have a response variable, we will only rely on the supervised machine learning models in our analysis. Table.1 shows the different feature of the datasets such as number of instances and number of predictors in each dataset, datasets characteristics and the date that the datasets are created.

Table 1: Detailed Description of Cancer Datasets (Clark et al., 2013)

Dataset 1	Dataset Characteristic	Multivariate	Number of Instances	1318	Missing Values?	Yes
	Attribute Characteristic	Categorical Integer	Number of Attributes	14	Date Modified	09/30/217
	Associated Tasks	Classification Regression	Number of Predictors	7	Response Variable	Pathology benign/ malignant
Dataset 2	Dataset Characteristic	Multivariate	Number of Instances	72	Missing Values?	Yes
	Attribute Characteristic	Categorical Boolean Integer	Number of Attributes	38	Date Donated	07/14/2015
	Associated Tasks	Classification Regression Training	Number of Predictors	16	Response Variable	Censor (0 or 1)

In the dataset 1, both predictor and response variables are categorical, hence we hardcoded the predictors with numerical values for the ease of analysis. The values were assigned on a rating scale of 0 to 2 with an interval of 0.5, 0 being less severity and 2 being high. Table.2 shows the correlations coefficient between response variables and predictor variables for dataset 1. As it is represented in Table.2, owing to low variance, there are minimal correlations between most of the features.

Table 2: Correlations between response and predictor variables for dataset 1

	breast_density	abnormality id	mass shape	mass margins	assessment	pathology	subtlety
breast_density	1.000000	-0.163951	0.074470	0.143668	0.060066	-0.043675	-0.290062
abnormality id	-0.163951	1.000000	-0.120767	-0.158748	-0.073784	-0.020285	0.085900
mass shape	0.074470	-0.120767	1.000000	0.601927	0.326844	0.469475	-0.093377
mass margins	0.143668	-0.158748	0.601927	1.000000	0.374541	0.518060	-0.140368
assessment	0.060066	-0.073784	0.326844	0.374541	1.000000	0.492378	0.076325
pathology	-0.043675	-0.020285	0.469475	0.518060	0.492378	1.000000	0.031646
subtlety	-0.290062	0.085900	-0.093377	-0.140368	0.076325	0.031646	1.000000

For implementing the machine learning methods, first, we fitted and trained the data using linear regression, however, it gave minimal accuracy score of 44%. Mean square error was high. Later, using k Nearest Neighbor (KNN) classifier we modeled the data, holding pathology as the response variable. The highly correlated features were divided into training data as 70% and test as 30%. Initial accuracy mean was 76% at optimal k-value at 5. After fitting the model and running the analysis again the accuracy score was increased to 80%, which means our model can predict the outcomes of the pathology of a tumor (whether malignant or benign) 80% accuracy.

In the dataset 2, most of the predictors' features were numerical; however, the response variable "censor" is Boolean (0- non-recurrence of tumor post-surgery, 1- recurrence of tumor post-surgery). We preprocessed the data dropped the features that are irrelevant and explored the data by plotting histograms and pair plot for all the features. Owing to its high variance, there was a good correlation as shown in the heat map in the Figure 1.

[1] http://www.cancerimagingarchive.net

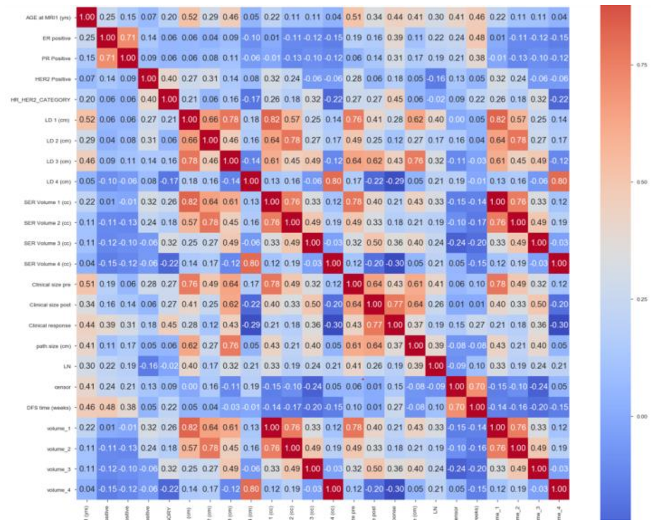


Figure 1: Heat map representation of correlations for data set 2

In the next step, we created three basic models and then optimized each model using Hyper-parameter Search technique. The model we used is Random Forest Classifier, KNN, and Support Vector Machine (SVM).
Random forest is capable of doing regression and classification, it can handle many features, and it is helpful for estimating which of the variables are important in the underlying data being modeled. We got 100% accuracy score, with no false positives or false negatives, which implies that the model over fitted the data since the number of instances, are very less (72). Then we applied KNN model, for which the accuracy score was 81%. Support Vector Machine model initially gave 70% accuracy, however, after fitting the model and training the data 92% accuracy score was achieved. The confusion matrix (Figure 2) shows no false positives (FP) but 1 false negative (FN) for the KNN model result.

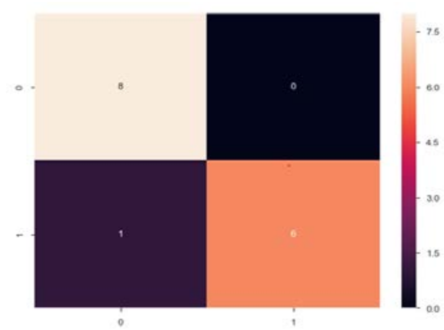


Figure 2: Confusion Matrix of predictive analysis for data set 2 using SVM method

4. Discussion on Literature Review

Given the growing importance of cancer prevention and the growing reliance on machine learning to make predictions, we believed it would be of interest to conduct a detailed review of published studies employing machine-learning methods in cancer prediction and prognosis. The intent is to identify key trends with respect to the types of machine learning methods being used, the types of training data being integrated, the kinds of endpoint predictions being made, the types of cancers being studied and the overall performance of these methods in predicting cancer susceptibility or patient outcomes. Interestingly, when referring to cancer prediction and prognosis most studies were concerned with three "predictive" foci or clinical endpoints: (1) the prediction of cancer susceptibility (i.e. risk assessment); (2) the prediction of cancer recurrence and (3) the prediction of cancer survivability. We also found that almost all predictions are made using just four types of input data: genomic data (SNPs, mutations, microarrays), proteomic data (specific protein biomarkers, 2D gel data, mass spectral analyses), clinical data (histology, tumor staging, tumor size, age, weight, risk behavior, etc.) or combinations of these three.
Depending on the type of data set, corresponding appropriate machine learning model would be selected for further analysis. For instance, deep learning and neural network methods are selected for datasets that have images (LeCun et al., 2015). For the data sets that are categorical and with high variance, classification machine models work better (Thrall et al., 2018).

From the case studies 1 and 4, evidently, machine-learning models can help radiologists read imaging studies from anywhere in the world at their home institution/office, bringing expert care to parts of the world that previously had limited expertise. Pathology has the opportunity to do the same with digital imaging and AI permits rapid and accurate local care. While still requiring evaluation within a normal surgical pathology workflow, deep learning has the opportunity to assist pathologists by improving the efficiency.
For the structured datasets that has clear predictor, variable supervised machine learning models are typically ideal. However, the above case studies show various supervised and unsupervised learning models implemented on breast cancer datasets for detection, diagnosis, and predictive analyses.
From our own data analysis, KNN gave decent accuracy score for data set1, where linear regression failed. For data set 2, SVM method appears to be the best fit in terms of predicting the results accurately. Especially, there are no FPs, which indicate that no patient would get unnecessary treatment because of the prediction mistake. In contrast, FN is the recurrence of cancer, which is predicted as non-recurrent, is very small in this case. In this case, there might be some time to reassess the patient in order to provide better treatment. Clearly, this not severe as the opposite situation.

Table 3: Comparison of accuracy scores obtained by various machine-learning models

	Model	Accuracy Score
Dataset 1	Linear Regression	43%
	KNN	80%
Dataset 2	Random Forest	100%
	KNN	81%
	SVM	92%

5. Limitations

In spite of the gaining popularity of employing AI and machine learning in detection and prediction of breast cancer, there are several fuzzy areas in determining the importance and robustness of machine learning models in detecting the breast cancer. There appear to be some practical problems as well in expecting to replace radiologists with machine learning algorithms. Some of them are as follow:

1. According to our research on literature review, we have noticed that among the more commonly noted problems was an imbalance of predictive events with parameters (too few events, too many parameters), overtraining, and a lack of external validation or testing.
2. Testing all basic and hybrid machine-learning methods on the test data was not possible. So we went with the popular methods that have been used for breast cancer prediction analyses in previous studies.
3. The existing papers in the literature review are broadly using a various number of machine learning models. This made it hard to compare each other's methodology, analysis, and their results in terms of robustness of their model. Furthermore, their data sets varied enormously in terms of the type of data dealt with. Some used images, some used blood sample slides in the pathological analysis, and some used discrete data, and so on. These differences made it hard to evaluate the strength of each model to determine their generalizability.
4. Small sized training sample, compared to data dimensionality can result in misclassifications while the estimators may produce unstable and biased models.

6. Conclusions

Success in machine learning is not always guaranteed. As with any method, a good understanding of the problem and an appreciation of the limitations of the data is important. So too is an understanding of the assumptions and limitations of the algorithms being applied. In our study, we found that choosing the most informative subset for a training model by means of feature selection methods could result in a robust model. If a machine learning experiment is properly designed, the learners correctly implemented and the results robustly validated, then one usually has a good chance at success. Techniques such as Machine Learning (ML) and Medical Expert Systems (MES) algorithms have added impetus to the use of artificial intelligence in detecting and diagnosing breast cancer. The future research can be concentrated on considering similar breast cancer datasets; either completely diagnostic or completely prognostic which may be more successful in determining robust learning model.

References

Bejnordi, B. E., Veta, M., van Diest, P. J., van Ginneken, B., Karssemeijer, N., Litjens, G., Balkenhol, M. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama, 318*(22), 2199-2210.

Braunstein, M. L. (2016). *Practitioner's Guide to Health Informatics*: Springer.

Cheng, J.-Z., Ni, D., Chou, Y.-H., Qin, J., Tiu, C.-M., Chang, Y.-C., . . . Chen, C.-M. (2016). Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Scientific reports, 6*, 24454.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., . . . Pringle, M. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging, 26*(6), 1045-1057.

Dilsizian, S. E., & Siegel, E. L. (2014). Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Current cardiology reports, 16*(1), 441.

Himes, D. O., Root, A. E., Gammon, A., & Luthy, K. E. (2016). Breast cancer risk assessment: calculating lifetime risk using the Tyrer-Cuzick model. *The Journal for Nurse Practitioners, 12*(9), 581-592.

Jerez, José M, Ignacio Molina, Pedro J García-Laencina, Emilio Alba, Nuria Ribelles, Miguel Martín, and Leonardo Franco. 2010. 'Missing data imputation using statistical and machine learning methods in a real breast cancer problem', Artificial intelligence in medicine, 50: 105-15.

Jupp, P. W. (2018). A complex systems approach to cancer prevention. Medical hypotheses, 112, 18-23.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436.

Lee, A., Cunningham, A., Kuchenbaecker, K., Mavaddat, N., Easton, D., & Antoniou, A. (2014). BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. *British journal of cancer, 110*(2), 535.

McGuire, S. (2016). World cancer report 2014. Geneva, Switzerland: World Health Organization, international agency for research on cancer, WHO Press, 2015. *Advances in Nutrition: An International Review Journal, 7*(2), 418-419.

Oyelade, O., Obiniyi, A., Junaidu, S., & Kana, A. (2017). *A modified select and test (ST) algorithm for medical diagnosis in an ad-hoc network environment.* Paper presented at the Electro-Technology for National Development (NIGERCON), 2017 IEEE 3rd International Conference on.

Purdy, M., & Daugherty, P. Why Artificial Intelligence is the Future of Growth.–Accenture, 2016, 27 p.

Ritchie, Marylyn D, Lance W Hahn, Nady Roodi, L Renee Bailey, William D Dupont, Fritz F Parl, and Jason H Moore. 2001. 'Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer', The American Journal of Human Genetics, 69: 138-47.

Saslow, D., Boetes, C., Burke, W., Harms, S., Leach, M. O., Lehman, C. D., . . . Sener, S. (2007). American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA: a cancer journal for clinicians, 57*(2), 75-89.

Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. *CA: a cancer journal for clinicians, 68*(1), 7-30.

Thrall, J. H., Li, X., Li, Q., Cruz, C., Do, S., Dreyer, K., & Brink, J. (2018). Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *Journal of the American College of Radiology, 15*(3), 504-508.

Tilanus-Linthorst, M., Bartels, C., Obdeijn, A., & Oudkerk, M. (2000). Earlier detection of breast cancer by surveillance of women at familial risk. *European Journal of Cancer, 36*(4), 514-519.

Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., & Li, L. (2016). Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Scientific reports, 6*, 27327.

Wright, J. D., Pawar, N., Gonzalez, J. S., Lewin, S. N., Burke, W. M., Simpson, L. L., . . . Herzog, T. J. (2011). Scientific evidence underlying the American College of Obstetricians and Gynecologists' practice bulletins. *Obstetrics & Gynecology, 118*(3), 505-512.

Heart Rate Variability of healthy individuals depends on sex and age: a complexity approach

Ana Leonor Rivera,[1,2,*] Bruno Estañol,[2,3] Ruben Fossion,[1,2]
Juan Claudio Toledo-Roy,[1,2] Juan Antonio López-Rivera,[2,4]
José Luis Gómez,[2,4] & Alejandro Frank[1,2,5]

[1]Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México,
Circuito Exterior S/N, Ciudad Universitaria, 04510 Coyoacan, Mexico City, Mexico.
[2]Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México.
[3]Laboratorio de Neurosiologia Clinica, Departamento de Neurologia,
Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico.
[4]Facultad de Ciencias, Universidad Nacional Autónoma de México,
[5]El Colegio Nacional, México,

[*]To whom correspondence should be addressed; E-mail: ana.rivera@nucleares.unam.mx.

Heart rate variability depends not only on age, but also on gender for young healthy subjects. Time series analysis in the time and frequency domains of subjects watching Disney's movie Fantasia reveals statistically significant differences for young men and women, which are lost for elderly. The heart rate of young women has less variability and is more rigid compared to that of young men. The statistical distributions of the heart rate of young men have a larger variability, are asymmetric and platykurtic; those of young women exhibit less variability and are more symmetrical; finally, those for older subjects are more Gaussian both for men as for women. These gender differences may be due to hormonal modulation of the cardiac system. However, the entropy of the time series is similar for young people independently of sex, and is statistically different for old and young subjects. This is related to the loss of non-linear dynamics of the heart rate variability.

Introduction

Human health is determined by the functioning of its various organs and the interactions between them through a homeostatic balance between robustness and adaptability to the changes in the environment (*1, 2*). It has been shown that it is possible to find biomarkers based on non-invasive physiological measured parameters that characterize the healthy state (*3, 4*), and that these biomarkers could help as early-warnings, auxiliary to the clinical diagnosis of different diseases, in particular Type 2 Diabetes Mellitus (*5, 6*), and other autonomic disorders (*7*). Although there are studies of the differences due to sex as well as age in cardiac rhythm variability of healthy humans (*8–11*) there are still no conclusive results in this regard.

Heart rate variability (HRV) reflects not only the functioning of the heart and the circulatory system but also their relation to the respiratory and central nervous systems. The autonomic nervous system modulates the cardiac cycle by central (respiratory and vaso-motor centers) and peripheral oscillations (due to respiratory and arterial movements) that are altered both with disease and age (*12, 13*). Due to this, there is an increasing interest in the analysis of physiological time series in the frequency and time domains to characterize the variability of the cardiac rhythm.

In this work we look for HRV biomarkers that depend on age and sex. For this we present a time-domain analysis using the statistical moments of the distribution, the autocorrelation function, Poincaré plots, as well as the spectral (frequency-domain) analysis of RR intervals of healthy subjects .

Study population

Here we have extracted the heartbeat RR intervals from the electrocardiograms (ECG) (*14*) of the Fantasia database of Physionet, compiled by the Massachusetts Institute of Technology and the Beth Hospital of Israel, and its freely available at http://www.physionet.org/ (*10*). This database contains series of groups of 7 young women, 6 young men, 8 elderly women, and 9 elderly men, who are all healthy subjects.

Results and discussion

Group averages of statistical parameters of HRV (standard deviation, coefficient of variation, skewness and kurtosis) are given in Table 1, and group averages of non-linear measures from Poincaré plots and entropy are given in Table 2.

As an example of the differences due to gender and age, Fig. 1 shows the time series, histograms, power spectral density (PSD), and Poincaré plots of the RR interval of a 21-year old man, a 21-year old woman, a 77-year old man, and a 75-year old woman in supine position watching the Fantasy movie (2 hours runtime). The PSD is the log-log plot of the square magnitude of the Fourier transform of the time series vs. frequency, while the Pincaré plot shows the $(i+1)^{\text{th}}$ RR interval vs. the i^{th} RR (*3*). Poincaré SD1 corresponds to the variance measured along the identity line, while SD2 the variance along the perpendicular direction to SD1. These parameters are indicators of the long-term and short-term variability, respectively.

Table 1: Group averages of statistical parameters of the heart rate variability biomarkers for the subjects of the Fantasia database.

Group	Age (years)	standard deviation (s)	coefficient of variation	skewness	kurtosis
Male young	25 ± 4	0.08 ± 0.05	0.05 ± 0.02	-1.4 ± 0.9	11 ± 9
Female young	27 ± 4	0.08 ± 0.03	0.10 ± 0.03*	-4.0 ± 0.9*	6 ± 10
Female old	74 ± 5	0.05 ± 0.01*	0.08 ± 0.02*	-1.3 ± 1.5	25 ± 30*
Male old	75 ± 5	0.11 ± 0.04*	0.05 ± 0.02	-0.3 ± 0.9*	1 ± 1*

* denotes statistically significant difference ($p < 0.05$) with respect to the the male young group.

Table 2: Group averages of the non-linear measures of the heart rate variability biomarkers for the subjects of the Fantasia database.

Group	Poincaré SD1	Poincaré SD2	Poincaré excentricity	Shannon entropy
Male young	53 ± 18	147 ± 53	0.93 ± 0.04	1.5 ± 0.2
Female young	33 ± 11*	97 ± 28*	0.93 ± 0.04	1.5 ± 0.1
Female old	28 ± 8*	63 ± 15*	0.9 ± 0.1	1.4 ± 0.2
Male old	30 ± 17*	87 ± 36*	0.92 ± 0.05	1.0 ± 0.2*

As seen in Fig. 1, the heart rate of a young man has a high non-symmetric variability in its histogram, with an almost invariant distribution of scale (PSD slope -1.02), and it has a random component larger than that of a young woman (when comparing their Poincaré diagrams). It is important to note that cardiac rhythms are also more rigid as individuals age, as illustrated by the Poincaré plots for a young man and an old man (Fig. 1).

Considering all the individuals, the Poincaré SD1 and SD2 parameters reveal a large variability for young men, which are statistically different from the values of young women. Poincaré SD1 and SD2 are similar to those of elderly men who have a high short-range correlation (Table 2). It is also very interesting that Shannon's entropy distinguishes between sexes for elderly adults but not for young adults, perhaps reflecting that old men suffer a greater cardiac alteration than women (Table 2).

In the frequency domain there is no statistically significant difference between the groups.

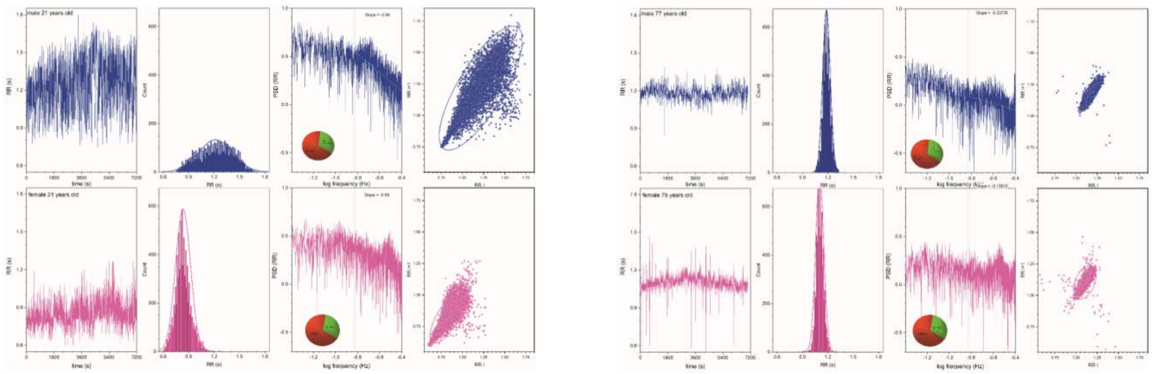


Figure 1: Plots from the RR records of a 21-year old man (top left), a 21-year old woman (bottom left), a 77-year old man (top right), and a 75-year old woman (bottom right) in supine position watching Fantasia movie. From left to right are the time series, histograms, Power Spectral Density (PSD) and Poincaré plots.

Conclusion

The distributions of RR intervals of young men, have a larger variability, are asymmetric and platykurtic; young women exhibit less variability and are more symmetrical; while elderly subjects display more Gaussian distributions. The heart rate of young women has less variability and is more rigid compared to that of young men, but for older subjects the gender dependence disappears. However, entropy for young men, young and elderly women have similar values, but they are statistically different from the entropy of eldrely men. This may be interpreted as a stronger loss of non-linear dynamics in the heart rate variability of elderly men compared to elderly women.

References

1. R. Fossion, J.P.J. Fossion, A.L. Rivera, O.A. Lecona, J.C. Toledo-Roy, K.P. García-Pelagio, L. García-Iglesias, B. Estañol, "Homeostasis from a time-series perspective: An intuitive interpretation of the variability of physiological variables", in *Quantitative Models for Microscopic to Macroscopic Biological Macromolecules and Tissues* ed. L. Olivares-Quiroz, O. Resendis-Antonio (pp. 87–109, Springer International Publishing, Cham, 2018).

2. R. Fossion, A.L. Rivera, Estañol, "A physicists view of homeostasis: How time series of continuous monitoring reflect the function of physiological variables in regulatory mechanisms". *Physiol Meas* **39**(8), 084007 (2018).

3. A.L. Rivera, B. Estañol, J.C. Toledo, R. Fossion, A. Frank, "Looking for biomarkers in physiological time series", in *Quantitative Models for Microscopic to Macroscopic Biological Macromolecules and Tissues* ed. L. Olivares-Quiroz, O. Resendis-Antonio (pp. 87–109, Springer International Publishing, Cham, 2018).

4. B. Estañol, A.L. Rivera, R. Martinez Memije, R. Fossion, F. Gómez, K. Bernal, ..., A. Frank, “From supine to standing: in vivo segregation of myogenic and baroreceptor vasoconstriction in humans.” *Physiol Rep* **4**(24), e13053 (2016).

5. A.L. Rivera, B. Estañol, R. Fossion, J.C. Toledo-Roy, J.A. Callejas-Rojas, J.A. Gien-López, ..., A. Frank, “Loss of breathing modulation of heart rate variability in patients with recent and long standing Diabetes Mellitus Type II.” *Plos One* **11**(11), e0165904 (2016).

6. A.L. Rivera, B. Estañol, H. Sentes-Madrid, R. Fossion, J.C. Toledo-Roy, J. Mendoza-Temis, ..., A. Frank, “Heart rate and systolic blood pressure variability in the time domain in patients with recent and long-standing Diabetes Mellitus.” *Plos One* **11**(2), e0148378 (2016).

7. J. Baker, J.M. Racosta, B. Balint, K. Kimpinski, “Utility of time and frequency domain parameters of Heart Rate Variability in the context of autonomic disorders characterized by orthostatic dysfunction”. *J Clinical Neurophysiol* **35**(2), 123–129 (2018).

8. J. Koenig, J.F. Thayer, “Sex differences in healthy human heart rate variability: A meta-analysis.” *Neurosci Biobehav Rev* **64**, 288–310 (2016).

9. A. Voss, A. Heitmann, R. Schroeder, A. Peters, S. Perz, “Short-term heart rate variabilityage dependence in healthy subjects.” *Physiol Measur* **33**(8), 1289–1312 (2012).

10. D.T. Schmitt, P.C. Ivanov, “Fractal scale-invariant and nonlinear properties of cardiac dynamics remain stable with advanced age: a new mechanistic picture of cardiac control in healthy elderly.” *Am J Physiol* **293**(5), R1923–R1937 (2007).

11. I.A. O’Brien, P. O’Hare, R.J. Corrall, “Heart rate variability in healthy subjects: effect of age and the derivation of normal ranges for tests of autonomic function.” *Br Heart J* **55**(4), 348–354 (1986).

12. M.N. Levy, “Sympathetic-parasympathetic interactions in the heart.” *Circ Res* **29**, 437–445 (1971).

13. TASK Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, “Heart rate variability. Standards of measurement, physiological interpretation, and clinical use.” *Eur Heart J* **17**, 354–381 (1996).

14. J.L. Gómez Reyes, A.L. Rivera, Electrocardiogramas: un enfoque desde la fisica de los sistemas complejos, Thesis UNAM, Mexico, 2018, online available at http://132.248.9.195/ptd2018/febrero/0770337/Index.html.

Acknowledgments

Partial Financial support for this work was provided by UNAM through grant PAPIIT/DGAPA IV100116 and CONACYT Frontiers grant 2016–01–2277.

PETRI
[Planet Earth Teaching & Research Initiative]
Media Portal: Ocean Plastic as a Complex System

Zann Gill

earthDECKS.org, project of The Ocean Foundation

PO Box 4001, Los Altos, CA USA 94024
zanngill@earthDECKS.org

Abstract. The Global Grand Challenge of PLASTIC on Earth today demands the skills of the complex systems community – the analytical tools, frameworks and understanding of Dynamic Social Networks.

earthDECKS [**DECKS** – **D**istributed, **E**volving, **C**ollaborative, **K**nowledge **S**ystem], under fiscal sponsorship of The Ocean Foundation, is developing PETRI [Planet Earth Teaching & Research Initiative] as a "petri dish" to study learning and collaborative problem-solving. Our first curriculum module is an app addressing ocean plastic as a complex systems problem. After beta testing, we selected a group of university professors to work with on customizing the app for university curriculum in several disciplines.

The app is inspired by Buckminster Fuller's concept for "World Game" (1961) as a learning game addressing systemic environmental challenges. World Game preceded the global internet, distributed computing, social networks, token offerings, and was fifty years ahead of the burgeoning "deep game movement" where players score high points for empathy and creative problem-solving, rather than for killing other players. The alternative currency movement offers a decentralized infrastructure for distributed apps manifesting collaborative intelligence, a way to revive Buckminster Fuller's concept for World Game, not merely as a game, but as a global problem-solving tool.

earthDECKS' first media portal, PLASTIC, tackles ocean plastic, a problem widely acknowledged and documented. Participants join a social network via a media portal that captures the unique profile of each user entering the system and, based on that user's profile, matches each user to resources, challenges, and other users with complementary interests. The DECKS in earthDECKS are clusters of "infocards" that players use to decide how to act. Each player receives an earthDECK, customized for that user's profile.

Keywords: blockchain, complex systems, learning games, distributed learning, ocean plastic, social networks, World Game

1. Introduction: Scope of the Problem

Our objective is to demonstrate a new type of problem-based learning, where case studies of past projects are replaced by current problems that students help to identify. earthDECKS is a learning ecosystem, with complementary online and offline interactions where learners can contribute to solving real world problems that require a diversity of skills and disciplines to communicate and collaborate.

earthDECKS.org chose to focus on plastic as a complex systems problem because the World Economic Forum prediction in 2016 that by 2050 our oceans will have more plastic than fish has raised global awareness, triggering massive mobilization to address a challenge with no simple solution.[1] Plastic is useful for medical devices, 3D printing, industrial and machine parts, lenses, greenhouses etc. so critical analysis of tradeoffs is complex.

The ocean plastics problem crosses disciplines, institutions, regions, conflicting economic interests and requires participation of people with diverse expertise and objectives, who do not normally work together – from innovators developing plastic substitutes to entrepreneurs developing plastic upcycling systems for applications, from road-building to housing and 3D printing for medical applications. Supply chain management professionals sourcing materials for products and packaging, environmentalists tracking plastic as it travels along waterways to the ocean, and marine biologists studying its impact on marine life and ecosystems need to understand each other's findings.[2] Plastic in the ocean comes from land, from plastic production, consumption and disposal.[3]

This paper introduces the earthDECKS App, PLASTIC, our first media portal for PETRI [Planet Earth Teaching & Research Initiative]. Our reason for choosing to address the need for knowledge exchange on this problem, and to build the framework to grow an Ocean Plastic Knowledge Ecosystem to communicate research and innovation stemmed not only from recognizing the need in this domain but also from realizing that many of the environmental crises we face are also complex systems problems.

Our app and back end knowledge platform support cross-disciplinary collaboration and innovation to address the many facets of the plastic problem. From corporate enterprise to philanthropy, from academia to government, earthDECKS' app is built on the principle of collaborative autonomy. As agents act with autonomy on a knowledge platform that offers guidance for more effective decisions, collaborative outcomes emerge.

1. 1. PLASTIC, an earthDECKS media portal for PETRI

Supply chain management professionals sourcing materials for products and packaging, environmentalists tracking plastic as it travels along waterways to the ocean, marine biologists studying its impact on marine life and ecosystems – all need to understand each other's findings. The PLASTIC supply chain can be impacted by interventions at any stage in this chain. We aim to raise awareness of the large window for innovation. Many plastic substitutes are being developed.

Our first objective is to enable a large network of people concerned about the plastic challenge to contribute skills and resources more effectively. The first step is to raise awareness, to engage a broad public audience in a social network tackling global problem solving challenges, starting with PLASTIC.

Our second objective is to improve critical thinking skills so that we can all decide how to work effectively on this hard problem.

Our third objective is to support game-changing innovation.

Our fourth objective is to create a petri dish in which to study how best to apply complex systems findings to motivate and incentivize dynamic social networks addressing global challenges.

Our fifth objective is to drive game-changing innovation in education, harnessing the power of new media to pioneer a next generation social network that demonstrates a new paradigm for MOOCs (Massive Open Online Courses) for active learning.

The earthDECKS app PLASTIC aims to harness the collaborative intelligence of users by matching them with other users in the system.

Collective intelligence defines the result obtained when effective algorithms are used to process input from a large number of anonymous human responders, all asked the same quantitative question. The focus is generally on predicting a quantitative outcome, producing a better-than-average prediction.

Collaborative intelligence describes a next generation social network where users are not anonymous. Unlike collective intelligence performers, who are anonymous, collaborative intelligence contributors have identifying signatures (original profile in the system) and footprints that evolve based on actions of the agents in the system.[4]

earthDECKS launched PETRI (Planet Earth Teaching and Research Initiative) as a way to explore how MOOCs can better support "citizen science" and learning through action. To engage the *collaborative intelligence* of diverse participants requires capacity to cluster and link related concepts, tag user profiles, and credit individual contributions. The focus is not on predicting the future, but on pattern recognition — measuring impact in the present, harnessing the unique capacity of online systems to capture user paths and time spent.

1.2. Historical Precedent: Buckminster Fuller's World Game

PETRI is inspired by Buckminster Fuller's concept for World Game. In the pre-Internet era, when massive multi-player online games did not exist, several hundred players assembled in university gymnasia to play World Game. A huge dymaxion map was laid out on the floor as the gameboard. A day-long improvisational theater experiment in collaborative problem-solving occurred. All players were given hard copy manuals, instructions and assigned roles, either as officials or as citizens of countries (with the number of citizen gamers proportional to each country's actual population). The players, from college-age to senior citizens, became engaged in this experiment. Some recounted the significance of what they realized through playing World Game — how it increased their understanding of our co-dependency on this planet.

For several decades this hard copy version of World Game was played around the world, and described as "a script … [for] a production ensemble ... a metaphor … successfully delineated by the consummate artist of his era – … Buckminster Fuller … with the design of his World Game." In May, 1968, the "World Game" was presented by Dr. Fuller to a White House sponsored conference, convened for this purpose, in Washington, D.C.; to the Muskie Committee to establish a select Senate Committee on Technology and the Human Environment in March, 1969; to the Joint National Meeting of the American Astronomical Society and the Operations Research Society of America in June, 1969; and to the United Nations Conference on Human Survival in May, 1970. After an intense flurry of interest, this Reality Game, conceived ahead of its time, was archived.[5]

Our experiment addresses the global problem of PLASTIC and aims to refocus the question from: "How do we analyze the impacts of plastics as a complex systems problem" to "How do we create a learning game where, by tracking actions, we analyze only what we can meaningfully analyze – the impact of learning in action." PETRI is a learning ecosystem where we conduct experiments that we can analyze. By tracking actions in PETRI, we create an evolving, self-improving learning system. Both the human learners and the AI system learn, and enable each other's learning. Human pattern recognition capacities work in tandem as an integrated system. The earthDECKS Ocean Network (eON) is our testbed of users, who are also enabling the system to evolve, so that our AI back end to learn from its users, and users can learn, both from each other and from recommendations from our AI system.

1.3. Historical Precedent: Wikipedia

earthDECKS, inspired by *Wikipedia,* is developing new generation of MOOCs with meta-tagging to network and cluster tags to similar and complementary tags (a hard problem that demands crowd-sourcing human pattern recognition); a rating, assessment, and filtering system, so that more useful lessons rise to the top; time-stamping, so that expiry and critical path dates and timeliness can be noted; authoring, such that Contributors are credited with a back-end system to track who contributed what; a system of search using not only key words but also parameter ranges; expertise and location of responders; date of incident and follow-up, clustering key words and other search criteria to filter results, not only by topic but according to other criteria; geotagging, so that where incidents occurred, and which responder networks were activated, is known to the system; incentives to motivate responders to report incidents, meta-tag and rate results, both their own contributions and those of others, particularly those that most relate to their own; capacity for Embedded Continual Assessment (ECA) to trace and report updates; transforming next generation social networks into problem-solving networks, with potential to scale and translate across varied requirements. This new type of MOOC (Massive Open Online Course) offers a unique way for students to navigate their own learning paths.[6]

1.4. PETRI: A New Model for MOOCs (Massive Open Online Courses)

PETRI [Planet Earth Teaching & Research Initiative] offers a new type of MOOC (Massive Open Online Course) with a unique way for students to navigate their own learning paths. After they watch the introductory film PLASTIC, clicking on topics of interest to them, they play a short CIQ (Collaborative IQ) game to see their customized earthDECK, responsive to their interests. Although they define their profiles in their initial signup, their choices in the system expand their profiles. Based on their profiles, PETRI makes recommendations, customized to each user's unique profile.

In the first iteration of PETRI for application in learning curriculum, every Story Contributor has the role of "Reporter" in an environmental news agency network. Every story reader is a "News Analyst," learning critical thinking skills by writing comments about the story. Players earn tokens by reading and writing stories, which receive critique and ratings (how token awards are determined) from other students and teachers. The best stories are posted and become part of PETRI's growing story collection. The backend analytics allow us to see which topics attract most interest and how interest grows.

Our downstream objective is not only a new paradigm for MOOCs, and for online learning in general, but also new paradigm for search that transforms social networks into problem-solving ecosystems. PETRI offers a better way to use media as portals to a knowledge platform to support convergent problem-solving.

2. The Urgency and Complexity of the Plastic Problem

This section describes four attributes that make the ocean plastic a problem suitable for developing cross-disciplinary problem-solving skills. First, the complexity of analyzing the health impacts of pervasive plastic pollution. Second, the cross-jurisdictional complexity of the problem. Third, the diversity of types of plastic and difficulties identifying and sorting waste plastic. And fourth, the difficulty assigning ownership of the problem. There are other windows on this problem. For example, the financial damage caused by plastic waste to tourism, recreation, and business has become inestimable. And translating impacts on human and animal health to loss of biodiversity and large scale ecosystem degeneration engages complex systems analysts with environmental scientists and designers of mitigation strategies.

Plastic is not only a complex systems problem; it is a problem where many uncertainties make collecting accurate data for analysis difficult. PETRI offers a way to convert the problem of analyzing past data into a problem in synthesis and convergent problem-solving, focusing on analysis of public perception of the plastic problem and how raising awareness can motivate action that is recorded and tracked on the earthDECKS knowledge platform. Although in 2018 there are an estimated 2.5 billion active online gamers worldwide, many of whom started young,[7] what is proposed is broader than a traditional game: PETRI engages players in serious real challenges where winning entails helping others to solve a real problem.

2.1. Health Impacts of Different Types of Plastics

Our first window on the complexity of the plastic problem is the health impacts of various types of plastic on unknowing consumers. Most of what we eat and drink contains plastics. Minute microplastics and fibres have been found in many products, from honey to sugar,[8] shellfish,[9] bottled and tap water,[10] beer,[11] processed foods, table salt[12] and soft drinks. 93% of adults tested in the US had detectable levels of the known carcinogenic chemical bisphenol A (plastic additive) in their urine.[13] 83% of samples of tap water tested in seven countries were found to contain plastic microfibers.[14] A study published in March 2018 revealed plastics contamination in more than 90% of bottled-water samples from 11 different brands. A German study found fibers and fragments in the 24 beer brands tested.[15] 259 water bottles from 27 lots across 11 brands, purchased from 19 locations in nine countries were tested to conclude that 94 percent of our drinking water and 93 percent of sampled bottled water worldwide contain plastic particles and chemicals.[16]

Beyond plastic in your drinking water and table salt, plastic is also found as Bisphenol A in mothers' breast milk and infant urine samples. BPA is a plasticizer and monomer used in many plastics and is also widely used as a building material and in epoxy applications.[17]

2.2. Cross-Jurisdictional Complexity of the Ocean Plastic Problem

Our second window on the complexity of the plastic problem is cross-jurisdictional complexity. The ocean plastics problem crosses disciplines, institutions, regions, nations, and conflicting economic interests, and requires participation of people with diverse expertise and objectives, who do not normally work together – from innovators developing plastic substitutes to entrepreneurs developing plastic upcycling systems for applications, from road-building to housing and 3D printing for medical applications.

Nature Scientific Reports published in March 2018 a paper by Laurent Lebreton, oceanographer with The Ocean Cleanup Foundation, reporting that the "Great Pacific Garbage Patch," once dubbed "the size of Texas" is now estimated at 4 to 16 times larger than past estimates.[18] This massive pile of swirling trash, located between California and Hawaii, was found to have the highest concentration of plastic ever recorded in one area. *The New York Times* wrote on March 22, 2018 that the Pacific Garbage Patch is growing exponentially, soon making even this revised estimate wrong. The area is roughly four times the size of California, composed of an estimated 1.8 trillion pieces of rubbish, weighing 87,000 tons. According to the Ocean Cleanup Foundation, sea turtles caught near the patch had eaten so much plastic that it made up around three-quarters of their diet.[19]

NBC reported: "The vast dump of plastic waste swirling in the Pacific Ocean is now bigger than France, Germany and Spain combined — far larger than previously feared — and is growing rapidly."[20] Widespread global news coverage suggests that the world is listening. But what can we do? This plastic tragedy impacts everyone but the garbage patch lies outside any national jurisdiction.

After-the-fact cleanup is not a long-term solution. Plastic that ends up in oceans needs to be stopped at its source. Much of the plastic in the ocean has traveled there via rivers and canals. A 2017 study from the International Union for the Conservation of Nature found that almost all of the plastic in the ocean comes from land, from rivers and runoff from highly populated coastal cities,[21] from plastic production, consumption and disposal. The rest comes from maritime activities, such as fishing and shipping. In tens of thousands of landfills across the globe, plastic leaches toxic chemicals into the groundwater, which flows into lakes, rivers and oceans; 91% of plastic is not recycled.[22]

Although Asia generates relatively little waste per person,[23] a recent study suggests that 86% of the plastic that ends up in our oceans is running through Asia's rivers, which account for more ocean plastic than all other continents combined.[24] 90% of the discarded plastic that flows from rivers into the world's oceans comes from 10 rivers. eight in Asia and two in Africa.[25] A global survey reported that the Indus River deposits more plastic than 47 other rivers combined.[26] And the Yangtze River is several times worse, carrying up to 1.5 million metric tons of plastic debris and micro-plastic into the Yellow Sea each year. In addition to the top two, the Yangtze and the Indus, the other eight are the Yellow, Hai, Pearl, Amur, Mekong, and Ganges in Asia, and the Niger and Nile in Africa.[27]

The UK Canal & River Trust, which manages inland waterways in England and Wales, warns that plastic is ruining Britain's rivers, putting water wildlife at risk because plastic floats on the surface of waterways rather than sinking to the bottom.[28]

Dutch researchers from the Ocean Plastic Foundation found that the Yangtze River's mouth, where the conduit meets the sea, had a plastic concentration of 4,137 particles per cubic meter—and contributed 20,000 tons (22,046 metric tons) of plastic every year to the oceans. In December 2016, two ships dumped more than 100 tons (110 metric tons) of waste, such as needles and plastic tubes, into the Yangtze River. Researchers at The Ocean Cleanup estimate that rivers carry an 1.15-2.41 million tons of plastic into the sea every year.[29]

According to research published by the Ocean Conservancy and the McKinsey Center for Business and Environment, up to 60% of ocean plastic trash comes from five countries: China, Indonesia, Philippines, Vietnam, and Thailand.[30]

But where were these plastics invented? A history of the development of plastic packaging lists most of the innovations as invented in the United States.[31] With the exception of China, which is Number One for both plastic production and plastic pollution, the other four countries on the top five high polluters list are not on the list of top producers.

We're blaming the receivers of our packaging for the garbage in our oceans.

Despite generating relatively little waste per person, compared to the consumer-oriented West, total waste generated by the continent of Asia adds up. China is sensitive to this hard-to-solve problem. China manufactures the most plastic products, around 74.7 metric tons in 2015, according to the 2016 report by Plastics Europe, followed by 49.8 metric tons from Canada, Mexico, and the US combined. China began charging consumers for plastic usage in 2008 to fight against pollution. The country's booming delivery industry, fueled by e-commerce, is often not recyclable. Last year China's delivery companies used 12 billion plastic bags.[32]

One policy change can have cascading complex system implications.[33] In the 1990s, China took in plastic for recycling, seeing an opportunity to reuse it profitably to make new tools and trinkets. Since China began reporting in 1992, China has imported 106 million tons of plastic waste, making up 45.1 per cent of all imports worldwide, processing garbage from 43 countries as well as its own.[34] China and

Hong Kong have collectively taken in 72.4 per cent of all plastic waste. Other countries found mass plastic relocation cheaper than recycling the plastic themselves. High-income countries are responsible for nearly 90 per cent of plastic exports since 1988, with the EU, North America and Japan the top exporters.

China's ban on importing plastic for recycling, stopped used plastics from flooding into the country. China had made a policy decision that this plastic inflow was a social and environmental hazard. Since China's ban against accepting the world's plastic recycling went into effect January 1, 2018, plastic recycling crises have been reported in Britain, Canada, Ireland, Germany, and several other European nations. Tons of rubbish still piles up in port cities like Hong Kong.[35]

2.3. The Diverse Types of Plastic Polluting the Ocean

Our third window on the complexity of the plastic problem is from the point of view of the recyclers, who must classify plastic. PET and polyethylene vie for first and second place in global production, both at more than 70 million tons. Estimates vary; it is difficult to pin down an accurate estimate. Polypropylene and polyvinyl chloride (PVC) vie for third and fourth place, both at more than 50 million tons. Polystyrene (PS), marketed under the trade name Styrofoam®, is number 5. Polyethylene, polypropylene, and polystyrene make up over 70% of the plastic used in medical devices, so the uses of these top plastic products are not all destructive. The most produced plastics in order are:

1. **polyethylene terephthalate (PET)** is the most common plastic, used in fibers for clothing, containers for liquids and foods, thermoforming for manufacturing, as a moisture barrier, and in combination with glass fiber for engineering resins. 30% of its production is used in plastic bottles. PET is commonly used in commercially sold water bottles, soft drink bottles, sports drink bottles and condiment bottles. Annual global production is estimated by 2020 at more than 70 million metric tons.
2. **polyethylene (PE)** vies for the title of "most common plastic." It is also used in packaging (plastic bags) and for plastic films, geomembranes, containers including bottles and grocery store bags.
3. **polypropylene (PP)** is the second most common plastic, used for packaging and labeling, textiles, plastic parts and reusable containers of various types, laboratory equipment, automotive components, and medical devices. Annual global production is c. 55 million metric tons.
4. **polyvinyl chloride (PVC)** vies with polypropylene for the title of "second most common plastic." PVC is used in bottles, cards (e.g. bank or membership cards); electrical cable insulation, flooring, imitation leather, inflatable products, non-food packaging, plumbing, phonograph records, signage, and as a rubber substitute. Annual global production is c. 54 million metric tons.
5. **polystyrene (PS)** – marketed under the trade name Styrofoam® is number five in global production at several million tons per year and is widely used in packaging (e.g. yogurt containers) and found in its foamed form as litter in the landscape and in the ocean, particularly the Pacific. Polystyrene is an aromatic hydrocarbon and is chemically very inert.
6. **polylactic acid (PLA)** – differs from other plastics in being derived from biomass rather than petroleum, so it biodegrades more rapidly than traditional plastic

materials and is increasingly used as a substitute in manufacturing plastic cups and in DIY 3D printing.

7. **polycarbonate (PC)** – strong, impact resistant, and transparent and so is used in greenhouses and in riot gear for police and also for reusable plastic containers, a source of BPA.
8. **acrylic** – highly transparent, scratch resistant and so is used in contact lenses, glasses, magnifying and other lenses.
9. **acetal (polyoxymethylene, POM)** – high tensile strength and high resistance to heat, abrasion, water, and chemical compounds. It is also creep resistant and has a low coefficient of friction so that it can be used in gears.
10. **nylon** – used in clothing, reinforcement in rubber material like car tires, rope, thread, in injection molded parts for vehicles and mechanical equipment, as a substitute for low strength metals in car engines because it is strong relative to other plastics, has high temperature resilience, and high chemical compatibility. Used in 3D printing.
11. **ABS (Acrylonitrile Butadiene Styrene)** – strong resistance to corrosive chemicals and physical impacts, is easy to machine, widely available, has a low melting temperature so easily used in injection molding, and so is the number one material in 3D printing, used for products like Lego blocks.[36]

To analyze the supply chains for these top eleven plastics requires finding out how much is produced, where, and how it is distributed for further manufacture into this range of products, how these products are distributed, their lifetime and disposal and finally the ability to classify disposed plastic for recycling and reuse.

2.4. Top Eight Plastic producers: Who owns the plastic pollution problem?

Another alternative is to start, not with the types of plastic, but with its producers.[37] Plastics manufacturing is the United States' third-largest industry, employing nearly one million people in plastics manufacturing or supply – 1.4 million plastics-related jobs, nearly 16,000 injection molding and plastic manufacturing facilities, producing nearly 113 billion pounds of plastics and resins (2017).[38]

Most plastic production does not occur in the countries blamed for polluting our oceans.[39] Only China is Number One for both plastic production and plastic pollution. The only other country, besides China, that is on both the list of top plastic producers and also on the list of top plastic polluters is the United States (Number 3 Producer, Number 20 Polluter). Imagine if China and the United States collaborated on the global challenge to reduce unnecessary plastic production and pollution. Here, in order, is the list of the Top Eight plastic producers:

1. China *$18 billion. 27.2% of Total Exports.*[40]
2. Germany *$7.8 billion. 11.8% of Total Exports.*[41]
3. United States *$6.3 billion. 9.5% of Total Exports.*[42]
4. Italy *$2.6 billion. 3.9% of Total Exports.*
5. France *$2.5 billion. 3.5% of Total Exports.*
6. Taiwan *$1.9 billion. 2.9% of Total Exports.*
7. Hong Kong *$1.8 billion. 2.8% of Total Exports.*
8. Mexico *$1.7 billion. 2.6% of Total Exports.*[43]

3. Ocean Plastic: Knowledge Ecosystem Promoting Innovation

Our learning app, knowledge platform and associated social network for action link a new form of personalized learning, described in Section 1, to an urgent problem, described in Section 2. The objective is to foster an ecosystem that promotes innovation, as described in this section.

3.1. Innovation to Address the Plastic Problem

Moving on from analysis starting with plastic health impacts, jurisdictional issues, plastic types, or plastic production capacity, suppose that we table the analysis challenge to focus on design synthesis — how the plastic problem has inspired innovation. Adidas launched their third collection of shoes made from upcycled ocean plastic.[44] Loop Industries, the Terrebonne, Canada-based waste recycling technology company, partnered with Paris-based Evian to create Evian's plastic bottles out of 100 percent recycled plastic (by 2025).[45]

Plastic substitutes from natural materials is a new field in materials science. Mycelia are a complex system of tubular cells that function as the mushroom's sense organs to sense the surrounding environment, absorb nutrients as food, so the fungus can grow and survive. Mycelia are the key component in a new kind of fungus-based packing material, a substitute for polystyrene (styrofoam®), resulting from university research and industry collaboration.[46]

UBQ is an Israeli company with a patented process to convert household trash, diverting waste from landfills into reusable bio-based plastic. For every ton of material produced, UBQ says that it prevents between three and 30 tons of CO_2 from being created by keeping waste out of landfills. UBQ says that its closely guarded patented process produces no carbon dioxide or toxic byproducts, and uses little energy and no water. In addition to plastic alternatives, innovators are also finding ways to turn plastic trash into a resource, as in the brilliant example of the Plastic Bank.[47]

3.2. The Plastic Bank

David Katz and Shaun Frankson, co-founders of the Plastic Bank, a startup from Vancouver, Canada, realized that we need new social structures to address the plastic trash problem. They asked, "Can we solve the problems of ocean plastic pollution and extreme poverty at the same time?" They aimed to make plastic waste too valuable to enter the ocean. They announced the concept of Social Plastic® and traveled to Peru and Colombia to learn from those who recycle for a living. By mid 2014, The Plastic Bank successfully inspired a million people to join the Social Plastic® movement, with supporters asking companies to use #SocialPlastic.[48]

The Plastic Bank now has a global chain of pop-up stores where everything from school tuition to cooking fuel is available to buy in exchange for plastic garbage, which is then sorted, shredded and recycled. The Plastic Bank is closing a loop in the circular economy.

In developing countries, about 80% of plastic refuse comes from areas with high levels of poverty and no effective waste management systems. IBM has applied its blockchain technology to support the Plastic Bank. With IBM Customer Cognition

Foundry the Plastic Bank developed a blockchain solution to overcome the risk and mistrust of a cash-based system. Through a blockchain reward system people could earn and spend Plastic Bank digital tokens, collecting enough plastic refuse to provide for their families. The plastic they bring to established recycling centers is exchanged for digital tokens they can use to buy goods: food, water, phone-charging credits etc. Anyone can exchange digital tokens for items in the store.[49]

3.3. Plastic Roads

Landfilling and incinerating plastic are both problematic methods of managing plastic waste. Plastics in landfills can leach pollutants into the surrounding soil; incinerating creates gaseous pollutants, such as carbon dioxide. Dumping plastic in the ocean destroys the ocean ecosystem on which all life depends.

The supply chain challenge posed by the plastic problem is addressed by one big, simple idea. The big idea was not a technology innovation; it was a simple breakthrough to "think different" – to use recycled plastic, rather than virgin plastic, enabling

1) Local recycling. Implementation of plastics in roads opens a new option for recycling post-consumer plastics near their point of origin, saving in transport and processing. Solving this supply chain problem had an added bonus:
2) Better Roads at lower cost. Civil engineers proposed that substituting plastic for bitumen in asphalt would produce stronger roads, but the original idea assumed virgin plastic, which was too expensive.

This big idea to use recycled plastic spawned a Global Innovation System to recycle plastic into roads, with leadership from India to the Netherlands, Scotland, and the US state of Texas, which is now having a global ripple effect. Plastic-bitumen composite roads wear better than standard asphalt concrete roads, absorb sound better, absorb water less, are more flexible, less prone to rutting and need less maintenance to remain smooth.

In the UK Engineer Toby McCartney proposed substituting recycled plastic pellets for the 10% bitumen (a petroleum product) now used in asphalt road mix and launched England's largest crowd-funding campaign, installing roads with their new product in 86 locations across London, achieving 150% of the investment goal in just 14 days.[50]

Chennai was an early adopter of plastic roads. Jambulingam Street in Chennai, Tamil Nadu was one of India's first plastic roads. The idea about 15 years ago to bury plastic waste beneath roads to remove plastic from landfills preceded broad realization that recycled plastic could be used to make better roads. Over time polymer roads proved durable. The 2015 floods that destroyed Chennai's roads raised awareness of the link between disaster recovery and innovation. The Mayor of Chennai announced that the plastic road project will move forward because plastic roads survive monsoon rains better. In November 2015, the Indian government announced that plastic roads would be the default method of construction for most city streets, part of a multibillion-dollar overhaul of the country's roads and highways. Today, there are more than 21,000 miles of plastic road in India, more than half in the southern state of Tamil Nadu. Most are rural roads, but that is rapidly changing.[51]

On June 22, 2018 the Maharashtra State Government legislated use of plastic waste to replace bitumen in asphalt for roads. Mumbai will now officially build and repair its roads using recycled plastic, rather than petroleum-based bitumen.[52]

The Netherlands is below sea level, innovating to address predicted sea level rise. In the Netherlands the Plastic Road Joint Venture is leading innovation in developing prefabricated modular plastic roads that can also serve as flood drainage systems. This innovation can be adopted in the Texas initiative, informed by recent experience of severe storms.[53]

3.4. Plastic Roads: Analysis of Pros & Cons

Plastic roads have many pros and very few cons (concerns):

Pro #1. Local upcycling of plastic trash for use near its disposal means that plastic trash in Mumbai can be used for plastic roads in Mumbai, plastic trash in Texas for roads in Texas etc. Using plastic trash in road construction reduces the amount of asphalt used, since plastic is used in a ratio of 8%, reducing use of petroleum-based bitumen in asphalt and also reducing plastics in landfills.

Pro #2. Flexible upcycling since plastic-bitumen composite roads can use a range of plastics, increasing the reuse of plastic. Plastic roads have more upcycling flexibility than most other uses. Most plastic waste is not recycled because it is usually mixed with different types of plastic and non-plastic (e.g. paper labels) The segregation process is labor-intensive with no easy solution. Plastic roads offer an easier solution.

Pro #3. Less costly. Using recycled, post-consumer plastics costs less than using bitumen, which saves resources, since asphalt is responsible for 2% of global carbon emissions and asphalt concrete requires petroleum.

Pro #4. More durable with better performance. Road surfaces remain smooth, cost less to maintain and absorb sound better.

Pro #5. Flood mitigation, heating, and power generation can be built into the design of plastic roads. Volkerwessels (part of the Netherlands Plastic Road Joint Venture) proposes building plastic roads with hollow space for wiring and drainage pipes. Heating and power generation can be incorporated into plastic roads. Heating prevents roads from freezing and can help evaporate water from the surface.

Pro #6. Modular prefabrication reduces onsite assembly time and road decommissioning. Fabrication of interlocking pieces, which can be quickly assembled or disassembled, reduces onsite road construction, and also reduces costs. Modular plastic pieces are easier and non-polluting onsite, unlike asphalt.

Pro #7. Equity and mobility in disadvantaged regions. Plastic roads make recycled plastic a more valuable commodity in disadvantaged countries where there is both an excess of plastic trash and the need for better roads. In this way, plastic roads promote equity.

Pro #8. Plastic roads can solve multiple problems simultaneously as shown in the pros listed above.

The cons (concerns) are technical issues to address to achieve the pros listed above.

Con #1. Plastic roads require using compatible plastics since melting together plastics of different types may lead to phase-separation and structural weaknesses, causing structural failure. Pro #2 above about plastic flexibility must be tempered with knowledge that all plastics cannot be combined and that research and testing is needed to determine the best mix for different environmental conditions.

Con #2. Abrasion of the surface of 100% plastic roads can release harmful plastic particulates that could exacerbate current microplastic pollution problem. This con needs to be carefully addressed in road design and construction, but it is counterbalanced by the many pros above.[54]

The City of Vancouver, Canada addresses these technical challenges through research and testing. An ambitious plan for Vancouver to be the greenest city in the world by 2020 includes early adoption of plastic road technology. The City is running many tests to determine the best mix.[55]

Conclusion

The PLASTIC Portal is our first media portal for PETRI. Users start by watching our short 5 minute film PLASTIC, our media portal to our backend knowledge platform and social network, which downstream can have other points of entry, other media portals focusing on other problems. Our interest is in studying how to optimize the performance of this platform to support dynamic social networks converging toward consilience about how to address complex systems problems. The three sections of this paper have outlined three key objectives of PETRI:

First, PETRI is defining a new emergent learning paradigm where the learner is in control as navigator, choosing her own path through information about a complex problem and contributing to the ecosystem according to her interests and motivation. The learner is engaged with other learners in a gamified social network that connects online learning to offline action in the real world.

Second, learners engage with an urgent real world problem that demands all hands on deck, contributing in different ways, and develops problem-solving skills through experience in action. The platform has templates for story contributors and rewards for participation.

Finally, learners not only grow their own learning ecosystem and networks with other learners. earthDECKS aims to promote innovation and problem-solving beyond the learning environment via trackable contributions in the real world.

Our media portal PLASTIC is just one puzzle piece in earthDECKS' planned cluster of portals focusing on cross-disciplinary grand challenges. Each portal will have its own viability, augmented by crosslinks to other portals.

References

[1] World Economic Forum Annual Meeting 2016. https://www.weforum.org/agenda/2017/01/more-plastic-in-sea-than-fish-3-strategies/

[2] Eunomia Research & Consulting Ltd. Plastics in the Marine Environment (2016).

[3] Jambeck, J. et al. Plastic waste inputs from land into ocean. Science. 347, 768-771 (2015).

[4] Gill, Zann. 2012. User-Driven Collaborative Intelligence: Social Networks as Crowdsourcing Ecosystems. ACM CHI (Computer Human Interaction). May 5 – 10, 2012. Austin Texas.

[5] Fuller, R. Buckminster. The World Game: Integrative Resource Utilization Planning Tool. Foreword by Thomas Broussard Turner. ix. January 1971. https://www.bfi.org/sites/default/files/attachments/literature_source/world_game_series_document1.pdf

[6] Gill, Zann. 2013. Wikipedia: Case Study of Innovation Harnessing Collaborative Intelligence. In: *The Experimental Nature of New Venture Creation: Capitalizing on Open Innovation 2.0*, edited by Martin Curley and Piero Formica. Dordrecht, Springer.

[7] https://www.wepc.com/news/video-game-statistics/

[8] Liebezeit, G., Liebezeit, E., Non-pollen particulates in honey and sugar. Journal of Food Additives & Contaminants: Part A. Volume 30, 2013 - Issue 12
pp 2136-2140. 28 Oct 2013. https://doi.org/10.1080/19440049.2013.843025
https://www.tandfonline.com/doi/abs/10.1080/19440049.2013.843025?journalCode=tfac20&

[9] Van Cauwenbergh, L., Janssen, C. Microplastics in bivalves cultured for human consumption. Environmental Pollution. Volume 193, October 2014, Pages 65-70. https://www.sciencedirect.com/science/article/pii/S0269749114002425

[10] Carrington, D. Plastic fibres found in tap water around the world, study reveals. The Guardian. September 5, 2017. https://www.theguardian.com/environment/2017/sep/06/plastic-fibres-found-tap-water-around-world-study-reveals
Readfearn, G. WHO launches health review after microplastics found in 90% of bottled water. The Guardian. March 14, 2018. https://www.theguardian.com/environment/2018/mar/15/microplastics-found-in-more-than-90-of-bottled-water-study-says

[11] Liebezeit, G., Liebezeit, E., Synthetic particles as contaminants in German beers. Journal of Food Additives & Contaminants: Part A. Volume 31, 2014 - Issue 9. https://www.tandfonline.com/doi/abs/10.1080/19440049.2014.945099?journalCode=tfac20

[12] Yang, D., Shi, H., Li, L., Li, J., Jabeen, K., and Kolandhasamy, P. Microplastic Pollution in Table Salts from China. Environ. Sci. Technol., 2015, 49 (22), pp 13622–13627. October 20, 2015. DOI: 10.1021/acs.est.5b03163 https://pubs.acs.org/doi/abs/10.1021/acs.est.5b03163?source=cen&

[13] NIH: https://www.niehs.nih.gov/health/topics/agents/sya-bpa/index.cfm

[14] op. cit. Carrington

[15] https://www.tandfonline.com/doi/abs/10.1080/19440049.2014.945099?journalCode=tfac20

[16] https://www.theguardian.com/environment/2018/mar/15/microplastics-found-in-more-than-90-of-bottled-water-study-says

[17] Mayo Clinic. https://www.mayoclinic.org/healthy-lifestyle/nutrition-and-healthy-eating/expert-answers/bpa/faq-20058331

[18] Lebreton, L., Slat, B., Ferrari, F., Sainte-Rose, B., Aitken, J., Marthouse, R., Hajbane, S., Cunsolo, S., Schwarz, A., Levivier, A., Noble, K., Debeljak, P., Maral, H., Schoeneich-Argent, R. Brambini & Reisser, J.. Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. *Nature Scientific Reports*, volume 8, Article number: 4666 (2018) https://www.nature.com/articles/s41598-018-22939-w

[19] The New York Times. March 22, 2018. https://www.nytimes.com/2018/03/22/climate/great-pacific-garbage-patch.html

[20] Galey, P, Hood, M., Pacific plastic dump far larger than feared.
March 22, 2018. https://phys.org/news/2018-03-pacific-plastic-dump-larger.html

[21] IUCN Report. 2017. https://www.iucn.org/resources/issues-briefs/marine-plastics

[22] Parker, L. Here's How Much Plastic is Littering the Earth. National Geographic. July 1, 2017. https://news.nationalgeographic.com/2017/07/plastic-produced-recycling-waste-ocean-trash-debris-environment/

[23] World Bank Report: What a Waste: A Global Review of Solid Waste Management. https://siteresources.worldbank.org/inturbandevelopment/Resources/336387-1334852610766/Chap3.pdf

[24] Lebreton, L., van der Zwet, J., Damsteeg, J.W., Slat, B., Andrady, A., Reiss, J. River plastic emissions to the world's oceans. Nature Communications. 7 Jun 2017. https://www.nature.com/articles/ncomms15611.epdf

[25] Patel, P. Stemming the Plastic Tide: 10 Rivers Contribute Most of the Plastic in the Oceans:The Yangtze alone pours up to an estimated 1.5 million metric tons into the Yellow Sea. February 1, 2018 https://www.scientificamerican.com/article/stemming-the-plastic-tide-10-rivers-contribute-most-of-the-plastic-in-the-oceans/

[26] Schmidt, C., Krauth, T., and Wagner, S. Export of Plastic Debris by Rivers into the Sea. Environmental Science & Technology 2017 51 (21), 12246-12253. DOI: 10.1021/acs.est.7b02368

[27] op.cit. Patel. P. Scientific American.

[28] Diver, T. Plastic in Britain's rivers is serious threat to wildlife, trust warns. Telegraph UK. March 31, 2018. https://www.telegraph.co.uk/news/2018/03/31/plastic-britains-rivers-serious-threat-wildlife-trust-warns/

[29] op. cit. The Ocean Cleanup. Scientific Reports. https://www.theoceancleanup.com/scientific-publications/ Lebreton, L. et al.

[30] McKinsey & Co. Stemming the Tide: Land-based Strategies for a Plastic Free Ocean. Summary: https://www.mckinsey.com/business-functions/sustainability-and-resource-productivity/our-insights/saving-the-ocean-from-plastic-waste/

[31] Plastic Packaging History: Innovations Through the Decades. Updated: January 5th, 2018. https://www.plasticsmakeitpossible.com/about-plastics/history-of-plastics/plastic-innovations-in-packaging-through-the-decades/

[32] Cheng Yu. China Daily. Food delivery firms push green. http://www.chinadaily.com.cn/cndy/2018-10/25/content_37133007.htm. Oct. 25, 2018

[33] United Nations Environment Programme. Marine Plastic Debris and Macroplastics – Global Lessons and Research to Inspire Action and Guide Policy Change (2016).

[34] Ellen Airhart. China Won't Solve the World's Plastics Problem Any More. *Wired Magazine*. June 20, 2018. https://www.wired.com/story/china-wont-solve-the-worlds-plastics-problem-any-more/

[35] Parker, Laura and Elliott, Kennedy. Plastic Recycling is Broken. National Geographic. June 20, 2018. https://news.nationalgeographic.com/2018/06/china-plastic-recycling-ban-solutions-science-environment/

[36] The Eleven Most Important Types of Plastic. Creative Mechanisms Staff. July 21, 2016 https://www.creativemechanisms.com/blog/eleven-most-important-plastics

[37] Top 10 Largest Plastic Producing Companies. https://www.plastics-technology.com/articles/top-largest-plastic-producing-companies

[38] The US Plastics Industry Numbers. http://indianabottle.com/u-s-plastics-industry-numbers/

[39] Nordrum, A. China, India, US Among World's Worst Marine Polluters: Country-By-Country Breakdown of Plastics in Oceans. International Business Times. 02/12/15 https://www.ib-times.com/china-india-us-among-worlds-worst-marine-polluters-country-country-breakdown-plastics-1813476

[40] http://www.chinatoday.com/china_suppliers/china_plastics.htm

[41]http://www.gtai.de/GTAI/Navigation/EN/Invest/Industries/Materials-processing/plastics.html

[42] http://www.firstamericanplastic.com/blog/10-facts-about-us-plastics-industry

[43] Eight Countries that Export the Most Plastic in the World. https://www.insidermonkey.com/blog/8-countries-that-export-the-most-plastic-in-the-world-574764/

[44] http://www.businessinsider.com/adidas-ocean-plastic-sneakers-reveal-a-new-reality-2018-3

[45]http://www.waste360.com/plastics/loop-industries-partners-evian-create-100-recycled-bottles

[46] Ecovative Design is the leading company developing mycelium-based packaging. https://ecovativedesign.com/

[47]https://www.israel21c.org/converting-chicken-bones-and-diapers-into-chairs-and-pails/?fb_-comment_id=1527456580709545_1531352320319971#f30d8371cc8afe4

[48] Field, A. The Plastic Bank: Using Plastic Refuse to Create a Global Currency for the Poor. *Forbes*. Nov 29, 2017. https://www.forbes.com/sites/annefield/2017/11/29/the-plastic-bank-using-plastic-to-create-a-global-currency-for-the-poor/#5fc096297433

[49] https://www.ibm.com/case-studies/plastic-bank & https://www.ibm.com/blogs/systems/plas-tic-bank-deploys-blockchain-to-reduce-ocean-plastic/

[50] Plastic Roads Bibliography: http://www.earthdecks.net/biblio-plastic-roads/

[51] Sribala Subramanian. Plastic Roads: India's radical plan to bury its garbage beneath the streets. The Guardian. June 30, 2016. https://www.theguardian.com/sustainable-business/2016/jun/30/plastic-road-india-tar-plastic-transport-environment-pollution-waste

[52] Swapnil Rawal. Maharashtra govt makes it compulsory to use plastic waste to lay roads. Hindustan Times. June 22, 2018. https://www.hindustantimes.com/mumbai-news/maharashtra-govt-makes-it-compulsory-to-use-plastic-waste-to-lay-roads/story-ZC6j1Rr0xZsuX4aZHH-bxhP.html

[53] Zann Gill. Plastic Roads: Global Innovation System. earthDECKS. http://www.earthdeck-s.net/plastic-roads/

[54] Gawande, Amit P. Economics and Viability of Plastic Road: A Review. Journal of Current Chemical and Pharmaceutical Sciences. Sept. 20, 2013. https://www.tsijournals.com/articles/economics-and-viability-of-plastic-road--a-review.pdf

[55] Vancouver Greenest City Action Plan. https://vancouver.ca/green-vancouver/greenest-city-action-plan.aspx

STEM

ACADEMIC PRESS

INTERNATIONAL CONFERENCE ON COMPLEX SYSTEMS 2018

NEW ENGLAND COMPLEX SYSTEMS INSTITUTE

Alfredo J. Morales

Editor