Article

Citizen Science and Topology of Mind: Complexity, Computation and Criticality in Data-Driven Exploration of Open Complex Systems

Masatoshi Funabashi

Sony Computer Science Laboratories, inc., Takanawa muse bldg. 3F, 3-14-13, Higashi Gotanda, Shinagawa-ku, Tokyo 141-0022, Japan; masa_funabashi@csl.sony.co.jp; Tel.: 03-5448-4380; Fax: 03-5448-4273.

Abstract: Recently emerging data-driven citizen sciences need to harness an increasing amount of massive data with varying quality. This paper develops essential theoretical frameworks, example models and a general definition of complexity measure, and examines its computational complexity for an interactive data-driven citizen science within the context of guided self-organization. We first define a conceptual model that incorporates the quality of observation in terms of accuracy and reproducibility, ranging between subjectivity, inter-subjectivity, and objectivity. Next, we examine the database's algebraic and topological structure in relation to informational complexity measures, and evaluate its computational complexities with respect to an exhaustive optimization. Conjectures of criticality are obtained on the self-organizing processes of observation and dynamical model development. An example analysis is demonstrated with the use of a biodiversity assessment database, the process that inevitably involves human subjectivity for management within open complex systems.

Keywords: inter-subjective objectivity; complexity measure; computational complexity; criticality; citizen science; open complex system;

Introduction

Recent innovation of information and communication technologies (ICT) embedded in real environment is drastically changing the way society interacts with computation. This has been described as the fourth industrial revolution [1]. In particular, ubiquitous sensors and mobile communication tools have led to an increasing capacity of distributed and interactive environmental sensing. These technological supports bring in new effective methodologies to tackle complex self-organising behaviours in social-ecological systems that are difficult to understand with conventional modelling and simulation approaches (e.g., [2] [3]). Massive amounts of sparse and heterogenous data that are based on the internal observation from within various collective phenomena call for an extended analytical framework, ranging from objective measurements such as with sensors, and subjective data such as human evaluations and feedbacks.

Redefining a standard formalization of computation and its complexity that are associated with self-organised citizen science can raise multiple criteria for the evaluation of critical phenomena, spread over the dynamical process of observation, management, and knowledge formation in open complex systems [4] [5]. Self-organised criticality appears in various natural and social phenomena, often with scale-free statistical properties [6][7]. They manifest in the power law, which can be reduced to a simple combination of inherent stochastic processes [8], and whose realizations provide proxies of emergent functionality (e.g., [9] [10] [11]). The large fluctuation of the power law distributes the statistical complexity in multiple scales that cannot be represented by a simple mean value for predictive purposes. The sampling time series from a power-law distribution encounters intermittent shifts of the sample average due to the infinite variance of distribution, even with the upper-bounded power law in the real world, e.g., in the magnitude distribution of earthquakes. This situation addresses a statistical limit of prediction solely by the modelling and simulation of the phenomena, but also presents a positive reason to engage human elements as a practical solution in actual management, especially those involving semantic and cognitive judgements [12] [13]. On the technology side, machine learning models have long been attempting to optimize the prediction of unknown stochastic sources, implementing interactive estimation processes to exploit the hidden causal structure from temporal observation sequences (e.g., [14]). Modelling studies of guided self-organization have been recently explored with the implementation to robotics, simulated neural networks and networks of agents, etc. [15]. Although most of the achievement is discussed within the predictability of a confined experimental setting, a hybrid system with the synergy of human and computation elements always lies as a premise of real-world situation, which has been little exploited, except for some prototypical interfaces for the internet of things (e.g., [16]). For a cost-effective monitoring and control within restricted resources, guided criticality should be introduced to the user side of technology, in order to migrate and abstract decision making process from computation to human ability [3] [4] [17].

In particular, in solving global agenda such as sustainability goals, a comprehensive approach is required that should make use of the full potential of self-organisation in coupled social-ecological systems [5] [18] [19]. These efforts practically take on the engagement of citizens and multi-disciplinary stakeholders as important actors in the data acquisition and the implementation of an interactive management through guided self-organization, as a novel type of collective intelligence in the era of the fourth industrial revolution [3] [20] [21].

In facing the transition of data-driven citizen science towards the achievement of dynamical control in managing real-world open complex systems, this article raises fundamental theories and example models to support the discussion of complexity, computation, and criticality in its most possible general form. We formalize the basic objectives as follows, which are exploited in the subsequent sections with the corresponding numbers:

- **Section** 1: How can we formalize and treat the databases of varying quality from both machine and human observations, which range from subjective bias to objective fact? How can we set up scientific measures that should assure the compatibility with the principles of accuracy and reproducibility ?
- **Section** 2: How can we generalize the concept of complexity measures in application to the human-computer hybrid systems in citizen science?
- **Section** 3: What is the nature of computational complexities in actual data processing ?
- **Section** 4: What is the general condition to yield guided self-organization for cost-effective citizen science ?

Although these questions are universal in multiple industries, a common basis of understanding the problems and mutual development of ICT infrastructure are still isolated and developed independently in each sector. Throughout the exploration of these topics, this paper attempts to provide a common terminology and establish a theoretical basis for the realisation of a cost-effective citizen science in open complex systems situations. This is becoming increasingly important for solving transdisciplinary problems through the participation of multiple stakeholders in real world [5].

1. Inter-Subjective Objectivity Model

We first consider the expression of the quality of data ranging between human subjectivity and machine objectivity in the general form of database $\mathbb{X}$. As a premise, any information that can be represented in digital computing is compatible with the natural number theory. At the infinite limit of computational memory, the representation of the database extends to general sets on a real data type with countably infinite precision, which accepts the definition of σ-finite measure in a measure-theoretical formulation. We define the general form of arbitrary database $\mathbb{X}$ as follows:

$$\mathbb{X} = \mathbb{R}^n \times \mathbb{S}^m \ (n, m \in \mathbb{N}). \tag{1}$$

Where $\mathbb{R}$ is a real data type, $\mathbb{S}^m$ is the m sets $\{S_i\}_{i=1,2,\ldots,m}$ of arbitrary symbolic set $S_i = \{s_1, s_2, \ldots, s_{l_i}\}$, with the dimensions n, m and l_i as natural numbers $\mathbb{N}$ including 0. Any variable in this article takes the assumption that it can be stored in $\mathbb{X}$. For mathematical simplicity, we hereafter consider the real data type $\mathbb{R}$ as a real number. In practise, $\mathbb{R}^n$ describes the values of n real variables (such as time, spatial coordinates, probabilities, etc), and $\mathbb{S}^m$ represents m discrete sets of symbols (such as the name of variables, occurrence of discrete variables, text data, etc). Obviously, $\mathbb{S}^m \subseteq \mathbb{R}^m$ holds in mathematical simplification, but we separate the notations to distinguish between the quantitative and qualitative variable types.

1.1. Formalization of Subjectivity, Inter-Subjectivity, Subjective-Objective Unity and Objectivity

Digital data $\mathbb{X}$ from citizen science vary from subjective human perception to objective sensor measurement with a different degree of human-induced bias. Here, the subjectivity and objectivity matter because it influences the accuracy and reproducibility of data that is fundamental to establish scientific analysis. We formalize the nature of observation variables between the subjectivity, objectivity and these interactions as follows:

- **Subjectivity** is the quality of observation that is based on human perception without the substantial support of a machine.
- **Inter-Subjectivity** is the degree of commonality between the subjectivities of multiple subjects.
- **Objectivity** is the quality of observation that is based on a machine measurement whose consequence does not depend on the operator's will.
- **Subjective-Objective Unity** is the degree of commonality between the subjectivity and objectivity.
- **Inter-Subjective Objectivity** is the quality of observation that satisfies the coincidence of both inter-subjectivity and subjective-objective unity.

These follow basic concepts in philosophy and social science and are adapted to the situation of data analysis. The concept of subjectivity is commonly used in philosophy as the collection of the perceptions, experiences, expectations, personal or cultural understanding, and beliefs specific to a person, which influences, informs, and is biased towards people's judgments and evaluations. In contrast, the objectivity refers to a view of truth or reality which is free from any individual's influence [22]. The most simplistical form of inter-subjectivity in social science employed the term in the sense of having a shared definition of an object, or shared subjectivity [23].

The relations between these classifications are shown in Fig.1(**a**). For example, text data written by humans are subjective data whether the fact described is based on an objective phenomenon or not. Sensor logs are objective data, even measured on a human body such as heart rate that could be influenced by subjective thought. When multiple subjects give the same subjective evaluation, such as rating of web contents, the commonality augments the degree of inter-subjectivity, which is often adapted to the cloud-sourced data validation (e.g., [24] [25]). When a subjective evaluation coincided with an objective measurement, the commonality represents the degree of subjective-objective unity. A highly reproducible subjective-objective unity can provide on-site practical measurement in field science, typical in the biodiversity assessment and soil texture analysis (e.g., [25] [26]). This is because these plausible subjective-objective unity measures also coincide with high inter-subjectivity after sufficient training, which guarantees the accuracy of on-site application without each time confirming the accordance with objective measurement. When the methodology is highly established with respect to the accuracy and reproducibility, it belongs to the inter-subjective objectivity, where each subjective and objective measurement converges to the same result. The developmental process of reproducible subjective evaluations that converge with objective measurements is depicted in Fig.1(**b**). By training the subjective-objective unity of each human observer, their inter-subjectivity increases, and the commonality of measurement augments to become a self-organizing loop between the subjective-objective unity and inter-subjectivity by a mutual feedback to attain a higher degree of inter-subjective objectivity.

Note that in a philosophical generalization, such as phenomenology, all data are the derivatives of subjectivity, because a machine observation is also constructed on human perception in the establishment of measurement principle, construction of sensing devices and data processing workflows, and final interpretation. To avoid trivial argument that does not affect the reproducibility of the results, we adopt the standpoint that separates the subjectivity and objectivity with the degree of intervention to observation outcome between human and machine. We call this conceptual model the **inter-subjective objective model**.

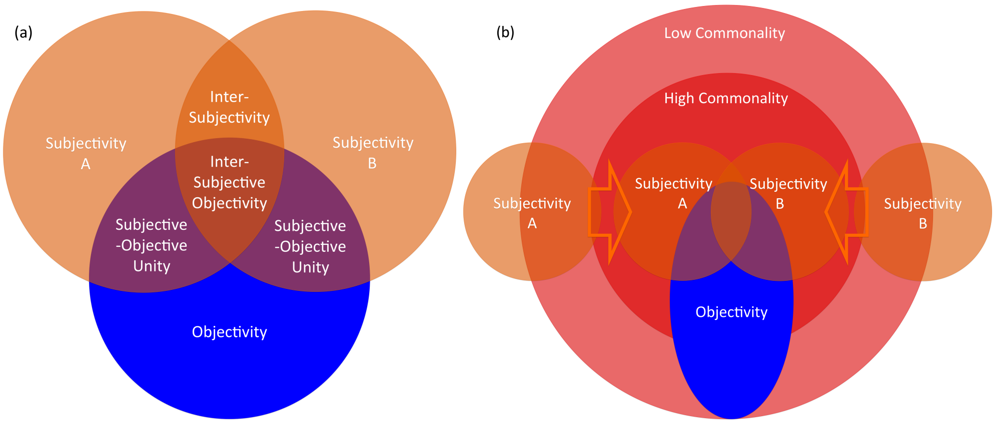

Figure 1. Schematic representation of the inter-subjective objectivity model. (**a**) Relations between two subjectivities namely A and B, objectivity, inter-subjectivity between A and B, subjective-objective unity for A and B, and inter-subjective objectivity are depicted as inclusion relations between each other set. (**b**) Development of inter-subjective objectivity as effective measurements of citizen science. As the inter-subjectivity increases along with the training of subjective-objective unity and inter-subjective feedbacks, the accuracy and reproducibility of measurement based on subjectivity can be assured by the convergence to inter-subjective objectivity.

1.2. Representative Model: Buoy-Anchor-Raft Model

In order to apply the inter-subjective objective model into quantitative framework of actual data processing, we develop a general example model with a more familiar and analogical terminology that are intuitively easier to understand: **Buoy-anchor-raft** model, as schematically expressed in Fig.2. The definition and correspondence to the inter-subjective objectivity model are given as follows:

- **Buoy** refers to subjective data that fluctuates on the sea surface representing subjectivity. Buoy can provide subjective estimates of an observation object lying on the objective sea floor, but the observation is biased by subjective fluctuations.
- **Anchor** refers to objective data that is fixed on the sea floor representing objectivity, without the influence from the subjective sea surface. Anchors can be connected to buoys, which provide the evaluation of subjective fluctuation with respect to objective machine measurements.
- **Raft** represents the relationship between buoys, and refers to inter-subjectivity of data without reference to anchors. A buoy can evaluate another buoy using relative difference of fluctuation on a subjective sea surface, and the overall commonality between buoys is represented as the raft. Nevertheless, it is based on an internal observation between buoys without an objective system of units, and is therefore susceptible to a global drift of collective standard.
- **Buoy-Anchor** connection rope defines the degree of subjective-objective unity. As a buoy's movement is more controlled by its anchor, higher subjective-objective unity is assured.
- **Raft-Anchor** connection ropes define the degree of inter-subjective objectivity. In addition to the commonality between buoys represented as a raft, the effects of the global drift from subjective sea surface could be controlled with anchors within a plausible range of error with respect to the objective sea floor.

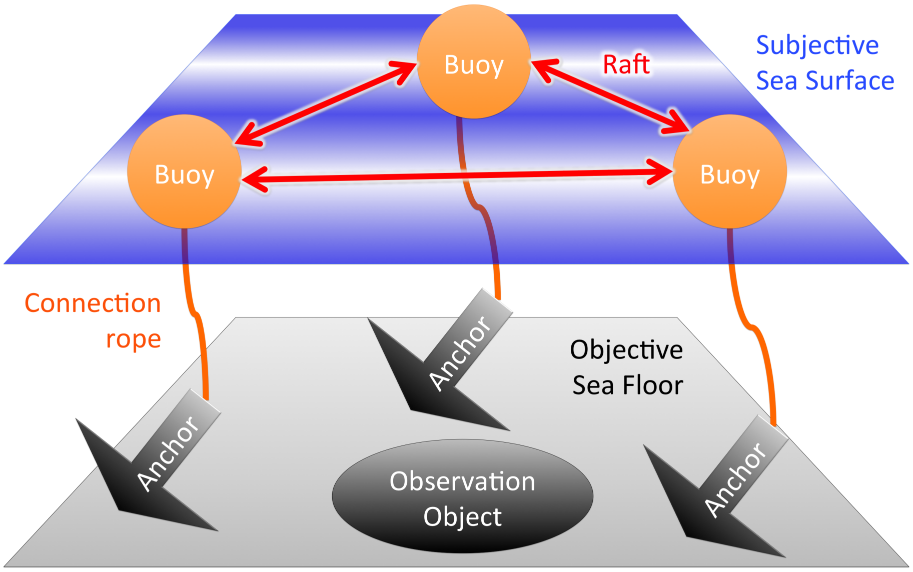

Figure 2. Schematic representation of **buoy-anchor-raft** model. **Buoy**, **raft**, **anchor** and connection rope refer to subjectivity, inter-subjectivity, objectivity, and subjective-objective unity, respectively. Concrete examples in the real world are given in Table 1.

Concrete examples of the **buoy, anchor, raft** in various social systems and scientific domains are given in Tab.1. While inter-subjective objectivity is a conceptual framework that classifies the quality of observation, the **buoy**, **anchor**, and **raft** refer to actual constructs of databases implemented with ICT. The terms arose from the developmental process of management systems in open systems science [5], sharing the perspective with the transversal question of the grand challenge of AI research regarding the effective extraction of scientific knowledge out of heterogenous data of varying quality [27]. Without properly positioning subjective background of the study, it is often the case that established knowledge with large-scale experiments and statistical analyses reveals to be false in high-throughput, discovery-oriented researches, resulting in a null-field with statistically prevailing bias [28]. As shown in Table 1, conceptual problematics for the implementation of ICT in various fields can be mutually characterized with the use of the **buoy-anchor-raft** model. This means the ICT infrastructure can be applied and shared in a synergistic way across domains, which is beneficial, especially for open-source development advocated in complex systems science [21]. Recent development in the Application Programming Interface for big data integration has increased the support for this challenge, which calls for a general theoretical framework of information processing that the **buoy-anchor-raft** model can provide (e.g., [29]).

Table 1. Examples of **buoy**, **raft** and **anchor** in various social systems and scientific domains. Examples are not comprehensive but a partial list of typical data from the recently increasing public availability.

	Economy	Judiciary	Biodiversity record	Medical treatment
Buoy	Demand, satisfaction	Sense of justice, guilt	Visual identification of species	Pain, psychological state
Raft	Price, exchange rate	Law, court decision	Identification with voting	Diagnosis, prescription
Anchor	Goods abundance	Evidential matter	DNA sequences	Physiological markers

We then consider a mathematical expression of the **buoy-anchor-raft** model in view of providing simplified idea of computation with respect to the evaluation of inter-subjective objectivity. Recently emerging contexts of citizen science make use of **buoys** as important information sources, in contrast to objective science such as traditional physics, which is usually self-contained with **anchors**. **Buoys** fluctuate with human subjectivity that is scientifically called bias. Suppose we cannot directly measure observation objects as **anchors**. This constraint does not necessarily arise from the observation principle but rather from the resource limitation: For example, a field evaluation of biodiversity mostly depends on human observation because massive DNA barcoding is too costly or even ineffective. So, the accuracy of **buoy** data should be evaluated with other **buoy-anchor** connections compatible with observation objects. By defining a buoy data $\mathbf{B} \subset \mathbb{X}$ and corresponding measurable anchor data $\mathbf{A} \subset \mathbb{X}$, a **buoy-anchor** connection $\mathbf{C}$ can be defined as an error function $\mathrm{erf}(\cdot)$ between $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{C} := \mathrm{erf}(\mathbf{B}, \mathbf{A}). \tag{2}$$

In case of n observation objects $\mathbf{A} = (a_1, a_2, \cdots, a_n) \in \mathbb{R}^n$ and $\mathbf{B} = (b_1, b_2, \cdots, b_n) \in \mathbb{R}^n$ for one observer, a typical example of **buoy-anchor** connection $c \in \mathbb{R}$ is given with the regularized mean squared error:

$$c = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{a_i - b_i}{a_i}\right)^2. \tag{3}$$

The regularization makes c accessible to the canonical evaluation of confidence interval such as t-test. As a generalization to m observers, let us describe

$$\mathbf{C} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}, \tag{4}$$

where

$$c_j = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{a_{ij} - b_{ij}}{a_{ij}}\right)^2, \ (j = 1, 2, \cdots, m), \tag{5}$$

given that

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{pmatrix}. \tag{6}$$

Next, we consider the **raft** model. In most social systems, the case-wise precise measurement of **anchors** is impossible and we call for the **raft** of common sense and other social feedbacks as a premise of plausible judgement. Consider m observers with somehow quantifiable opinions (**buoy**) on n observation objects. We define the **raft** matrix $\mathbf{R}$ as follows, as a generalization of **buoy** data to m observers and n observation objects:

$$\mathbf{R} = \begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{pmatrix}, \tag{7}$$

where the **raft** by definition refers to the commonality contained between these **buoys**. In a completely equal society where every observer's opinion is equally respected, we obtain the mean inter-subjective evaluation $\mathbf{E} = (e_1, \cdots, e_n)$ on n objects as follows:

$$\mathbf{E} := \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{pmatrix} \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix}. \tag{8}$$

Decision making based on the evaluation of **raft** can represent the community's mean quantifiable opinions, although it is not free from collective bias. It remains only within the framework of inter-subjectivity. For a better evaluation in terms of inter-subjective objectivity, we need to introduce a connection with **anchors**. Let us introduce a **buoy-anchor** connection **C** from Eq.(4), then an example of the inter-subjective objective evaluation $\mathbf{E}' = (e'_1, \cdots, e'_n)$ in the sense of **raft-anchor** connection can be given by:

$$\mathbf{E}' := \begin{pmatrix} e'_1 \\ \vdots \\ e'_n \end{pmatrix} \propto \begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{pmatrix} [-log(\mathbf{C})], \tag{9}$$

where

$$[-log(\mathbf{C})] = \begin{pmatrix} -log(c_1) \\ \vdots \\ -log(c_m) \end{pmatrix}. \tag{10}$$

This means that the error function of the **buoy-anchor** connection is reflected as an entropy that represents subjective-objective unity of each observer. The opinion of the observer with higher subjective-objective unity is weighted according to the informational scarcity of subjective errors. Such integrated evaluation incorporating the scoring system on observers' quality are one of the general solutions in web-based citizen science (e.g., [25]).

Note that the n objects of observation can also coincide with m observers themselves. As **C** can be independently obtained from **R**, it can also accept subjective objects of observation where direct **anchors** do not exist, such as psychological state or the quantification of qualia such as QFD [30] and pain scale [31]. In such cases, traditional methods only employ simple **raft** evaluation **E** without **anchors**, as formalized in Eq.(8). In contrast, with the **buoy-anchor-raft** model, it is possible to relate indirect **anchors** to other related objectively quantifiable variables, by expanding the database into a more comprehensive system. In either case, this model provides accessibility to the inter-subjective objective evaluation by properly defining the **buoy, anchor, raft** and its connections.

The correspondence between the **buoy-anchor-raft** model and computational variables developed in the following sections are listed in Tab. 2.

Table 2. Correspondence between **buoy-anchor-raft** model and computational variables in this article.

Section number	1.2	2.1	2.2	2.3	2.4	3	4
Buoy	**B**	$\mu(\cdot)$, $I(\cdot)$	$\mu_i(\cdot)$, $q_i(\cdot)$	Data contained in vertices V	Com. order I and II between N objects	Observations A, B, C, D, E	$P(\cdot)$, $P_a(\cdot)$, $P_s(\cdot)$, $P_l(\cdot)$, $P_o(\cdot)$, H'
Anchor	**A**						
Raft	**R, E**	$\mu'(\cdot,\cdot)$, $I_2(\cdot,\cdot)$	$\lambda_N(\cdot)$	Edge attribute of E	Com. order I and II b/w N observers, TDC, I-I and I-N res. dim.	$O(\cdot)$	H'_2, $D^m(\cdot:\cdot)$, $D^e(\cdot:\cdot)$
Buoy-anchor	**C**						
Raft-anchor	**E′**						

2. Complexity Measures

We consider the generalization of complexity measures with respect to essential information processing in citizen science, based on the inter-subjective objectivity model with **buoy-anchor-raft** constructs. The concept and definition of complexity vary according to the fields, such as algorithmic complexity, statistical complexity, biological complexity, etc. In this paper, we take a generalized definition of complexity measure as the projection from a system's variables to one-dimensional quantity, which is composed to express a distinctive characteristic of the system [32]. This includes classical indices mentioned with the context of complexity, as well as various forms of information expressed as numbers in ICT, such as feature dimensions of machine learning.

2.1. Complexity Measure and Search Function

We consider general forms of complexity defined on database $\mathbb{X}$ in relation to the search function. Complexity measures are widely studied in information theory, with the underlying principle to abstract a low-dimensional representative index of useful features for functional characterization of complex systems [32]. Usually, complexity measures defined on n real variables are the epimorphism to the one-dimensional real number line, $\mathbb{R}^n \mapsto \mathbb{R}$. The general complexity measure for citizen science is therefore the projection of the database to real value index, $\mathbb{X} \mapsto \mathbb{R}$, with the condition that this transformation will provide some utility for the management.

The importance of utility depends on the need for information retrieval in citizen science process, or the conditions that are practically used in a database search. Indeed, the search function is actually the retrieval of corresponding data set with respect to a given condition, such that

$$S_R[Q(x)] := \{x \in \mathbb{X} | Q(x)\}, \tag{11}$$

where S_R stands for the search result on database $\mathbb{X}$ with search query $Q(\cdot)$. For example, $Q(\cdot)$ is an if-then construct that can specify the value range of real variables, or the matching with specific symbolic sequence, which returns the corresponding data sets into S_R.

In order to perform computation such as the calculation of the **buoy-anchor-raft** model evaluation, the integral I of σ-finite measure μ on $\mathbb{X}$ with respect to the condition $Q(\cdot)$ can be defined as follows, with indicator function $\mathbf{1}(\cdot|Q(\cdot))$:

$$I(Q(x)) := \int_{\mathbb{X}} \mathbf{1}(x|Q(x))\mu(dx), \tag{12}$$

where

$$\mathbf{1}(x|Q(x)) := \begin{cases} 1 & if \quad x \in S_R[Q(x)], \\ 0 & if \quad x \notin S_R[Q(x)]. \end{cases} \tag{13}$$

In one-dimensional case, μ can represent either of **buoy** or **anchor**. If we define $\mu : \mathbb{X} \mapsto \mathbb{R}$ as the function of occurrence probability $p(\cdot)$ of $x \subset \mathbb{X}$, such as

$$\mu(x) = -p(x)log(p(x)), \tag{14}$$

then I coincides with entropy, one of the typical information theoretical complexity measures. μ can also include joint distribution, such that with μ':

$$\begin{aligned} \mu'(x,y) &= p(x,y)\log\frac{p(x,y)}{p(x)p(y)}, \\ x &\neq y, \\ x,y &\in \mathbb{X}, \end{aligned} \tag{15}$$

in which case the mutual information I_2,

$$I_2 := \int_{\mathbb{X}} \mathbf{1}(x,y|Q(x),Q(y))\mu'(dx,dy) \tag{16}$$

can incorporate **raft**, **buoy-anchor** and **raft-anchor** connections.

As a search query, $Q(x)$ provides a value of complexity measure I, we can also inversely use I to specify $S_R[Q(x)]$. We consider the invertible map $S_R^{-1} : \{x \in \mathbb{X}|I\} \to \{Q(x)\}$ that generates all possible queries $\{Q(x)\}$ which return the set of x associated with the given value of complexity measure I. For example, we can search the dataset with its entropy higher than a threshold I_c, by setting

$$\{Q(x)\} := S_R^{-1}\left[\left\{x \subset \mathbb{X} \middle| \int_x \mu(dx) > I_c\right\}\right]. \tag{17}$$

Nevertheless, complexity measures that specifically define an arbitrary $Q(x)$ are generally not given explicitly. In practice, we usually compare the performance of known complexity measures with respect to the ability to characterize the features we focus our analysis on. The general task is to invent a novel complexity measure that can exclusively separate patterns in $\mathbb{X}$, given implicitly as $Q(x)$. For that purpose, the following theorem holds:

Theorem 1. *For any search condition $Q(x)$, we can construct an exclusively selective complexity measure I' which can sort out effects from other variables, with the function $G(\cdot) : \mathbb{R} \mapsto \{Q(x)\}$, such that*

$$\begin{aligned} Q(x) &= S_R^{-1}\left[\{x \in \mathbb{X}|I'\}\right] = G(I'), & (18)\\ I' &= G^{-1}(Q(x)). & (19)\end{aligned}$$

The definition of invertibility of G follows that of S_R.

Proofs of the theorems are given in Appendix.

The intuitive geometric meaning of the inverse function relationship between complexity measures and search function is shown in Fig.3.

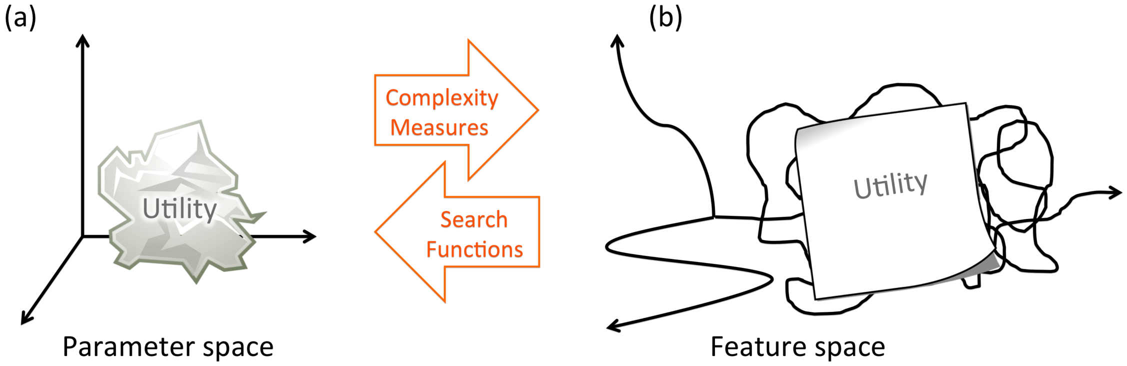

Figure 3. Schematic representation of complexity measures as non-linear feature space and search function as its inverse functions. (**a**) Utility characteristics of a complex system, or complexity measure in general terms, is expressed with a complex configuration in parameter space. Parameters can also represent other complexity measures. (**b**) Complexity measures transform parameter space into non-linear feature space, which provides easier interpretation by sorting the order of a given utility. The inverse functions of complexity measures therefore correspond to search functions with respect to the search condition on utility.

2.2. Observation Commonality as Complexity

Inter-subjective objectivity is based on the commonality among subjectivity, inter-subjectivity and objectivity. Essential computation is therefore the search for commonality between different observation datasets whether it be from humans or machines. We consider the observation commonality to be a complexity measure that conforms to inter-subjective objectivity, and analyze its general mathematical structure.

We consider σ-finite probabilistic measures μ_1, μ_2 on measurable database space $(\mathbb{X}, \mathcal{B})$, where $\mathcal{B}$ stands for Borel σ-algebra of $\mathbb{X}$. Then the convolution $*$ of μ_1 and μ_2 is defined as follows:

$$\mu_1 * \mu_2(s_i) := \sum_j \mu_2(s_j)\mu_1(s_{i-j}) \qquad \text{for } s_i \in \mathcal{B}(\mathbb{S}), \{s_i, s_j, s_{i-j}\} \in \mathbb{S}, \tag{20}$$

$$\mu_1 * \mu_2(\mathbf{x}) := \int_{\mathbb{R}} \mu_1(\mathbf{x}-y)\mu_2(dy) \quad \text{for } \mathbf{x} \in \mathcal{B}(\mathbb{R}),\ \mathbf{x}-y := \{x-y | x \in \mathbf{x}\}, \tag{21}$$

where $\mathcal{B}(\mathbb{S})$ and $\mathcal{B}(\mathbb{R})$ represent σ-algebra of $\mathbb{S} \subset \mathbb{X}$ and $\mathbb{R} \subset \mathbb{X}$, respectively.

Through appropriate variable transformation, the convolution of probability measures with real type variables (21) can be expressed as follows, as the probability of the sum of the variables [33]:

$$\mu_1 * \mu_2(\mathbf{x}) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}(x+y | x+y \in \mathbf{x}) \mu_1(dx)\mu_2(dy), \qquad \mathbf{x} \in \mathcal{B}(\mathbb{R}). \tag{22}$$

By choosing finite sets of $\mathbf{x}$ such as time period, geographic range, and other real type variable range, as well as symbols for $\{s_i\}$ such as name of observation object, one can define the commonality of observations as a part of the convolution of the probabilities from different observers. The observation μ_1 and μ_2 can be of any nature between subjectivity, inter-subjectivity, and objectivity.

We now consider the condition of valid observation with respect to the regularization of probability measure as follows, for a general number of observers $i \in \{1, \cdots, N\}$:

$$\int_{\mathbb{R}} \mu_i(dx) = 1. \tag{23}$$

This means, by expanding the scale of the real type variable to infinity, one can observe its occurrence with probability 1. The same formalization also applies to σ-finite measure on $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$, which is integrated in the formalization with $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Next, consider a confined variable range $r \subset \mathbb{R}$ with positive probability measure $\mu_i(r) > 0$. This range can be of any complex form as long as it supports positive measure. In a real situation, this can correspond to intermittent observation time interval, scattered geographical range, and other discrete range of the real type variable. We define the rate of observation q_i by observer i within variable range r as

$$q_i(r) := \int_{\mathbb{R}} \mathbf{1}(x | x \in r)\mu_i(dx) \leq 1, \tag{24}$$

which converges to (23) with $r \to \mathbb{R}$.

The commonality of observation between 2 observers i, j based on r is expressed as the following convolution confined to r:

$$\begin{aligned} \mu_i * \mu_j(r_2) &:= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}(x+y | x+y \in r_2; x, y \in r)\mu_i(dx)\mu_j(dy) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}(x+y | x, y \in r)\mu_i(dx)\mu_j(dy) \\ &= \int_r \int_r \mu_i(dx)\mu_j(dy), \\ r_2 &:= \{x_1 + x_2 | x_1, x_2 \in r\}, \end{aligned} \tag{25}$$

which also means taking the sum of joint distributions $\mu_i \cdot \mu_j$ between all smallest measurable events in r. The additional condition $x, y \in r$ in $\mathbf{1}(\cdot)$ limits the integral of each variable within r, which includes formal condition $x + y \in r_2$. The following generalization holds:

Theorem 2. *For N independent and valid observation $\mu_i(r) > 0$ $(i = 1, \cdots, N)$ on variable range $r \subset \mathbb{R}$, let*

$$\begin{aligned}
\lambda_N(r_N) &:= \mu_1 * \mu_2 * \cdots * \mu_i * \cdots * \mu_N(r_N) \\
&:= \int_{\mathbb{R}^N} \mathbf{1}\left(\Lambda \sum_{i=1}^{N} x_i \middle| \Lambda \sum_{i=1}^{N} x_i \in r_N; x_i \in r\right) \prod_{i}^{\{1,\cdots,N\}} \mu_i(dx_i) \\
&= \int_{\mathbb{R}^N} \mathbf{1}\left(\Lambda \sum_{i=1}^{N} x_i \middle| x_i \in r\right) \prod_{i}^{\{1,\cdots,N\}} \mu_i(dx_i), \qquad (26)\\
r_N &:= \left\{\Lambda \sum_{k=1}^{N} x_k \middle| x_k \in r\right\}, \\
\Lambda &:= \mathbb{R} \setminus \{0, \pm\infty\},
\end{aligned}$$

where the coefficient Λ *is a free parameter that remains invariant under the convolution. Then*

$$\lambda_N(r_N) = \prod_{i}^{N} q_i(r). \qquad (27)$$

This means that the N^{-1}-th power of multiple convolution $\lambda_N(r_N)$ represents the geometric mean of N independent valid observation rates. By choosing regularization factor Λ, r_N corresponds to the ensemble of possible mean values $\left(\Lambda = \dfrac{1}{N}\right)$, integrated values $(\Lambda = 1)$ and other weighted sum of N random samplings from r. The regularization parameter Λ can further be generalized to an arbitrary measurable function $\Lambda(\cdot)$ representing commonality characteristics, taking $\sum_{i=1}^{N} x_i$ as a variable.

With the use of the logarithmic scale, the information of $\lambda_N(r_N)$ is the sum of those with individual observation:

$$-log(\mu_1 * \mu_2 * \cdots * \mu_i * \cdots * \mu_N(r_N)) = \sum_{i}^{N}(-\log \mu_i(r)). \qquad (28)$$

As a similar property related to geometric mean, note that the following Young's inequality also holds:

$$|\Lambda| \cdot ||\mu_1 * \mu_2 * \cdots * \mu_i * \cdots \mu_N|| \leq \prod_{i}^{N} ||\mu_i||, \qquad (29)$$

where $||\cdot||$ denotes total variation. This assures us that the variation of the commonality remains within the order of the product of each observation's variation.

However, it is important to note that as a general property of convolution,

$$\lambda_N(r) \neq \prod_{i}^{N} q_i(r). \qquad (30)$$

The equality only holds in case $r \to \mathbb{R}$ or $\mu_i(r_N) = \mu_i(r)$ for $i = 1, \cdots, N$, without implication for the independence of observations. For the convolution on general subset $r_s \subseteq r_N$, the exact definition is given by

$$\lambda_N(r_s) \quad := \quad \int_{\mathbb{R}^N} \mathbf{1}\left(\Lambda \sum_{i=1}^{N} x_i \middle| \Lambda \sum_{i=1}^{N} x_i \in r_s; x_i \in r\right) \mu_1(dx_1)\cdots\mu_N(dx_N), \tag{31}$$

though it requires direct calculation without relevance to $q_i(r)$. In order to obtain fast computable form, the following asymptotical generalization holds:

Theorem 3. *As $N \to \infty$, for $r \subset \mathbb{R}$, $\mu_i(r) > 0$, $i = 1, \cdots, N$ and $r_s \subseteq r_N$, $\lambda_N(r_s)$ converges almost everywhere to the following:*

$$\lambda_N(r_s) \quad \to \quad \int_{r_s} \mathcal{N}(\Lambda\nu_N, \Lambda^2\sigma_N^2)m(dx) \times \prod_i^N q_i(r), \tag{32}$$

where $m(\cdot)$ is the Lebesgue measure on $\mathbb{R}$, and $\mathcal{N}(\nu_N, \sigma_N^2)$ represents the normal probability density distribution with mean value ν_N and variance σ_N^2 as follows:

$$\begin{aligned} \nu_N \quad &:= \quad \sum_{i=1}^{N} \int_{\mathbb{R}} \mathbf{1}\left(x | x \in r\right) x\mu_i(dx), \\ \sigma_N^2 \quad &:= \quad \sum_{i=1}^{N} \left(\int_{\mathbb{R}} \mathbf{1}\left(x | x \in r\right) x^2\mu_i(dx) - (\nu_N^2 - 2\nu_N) \right) \\ &\to \quad \sum_{i=1}^{N} \left(\int_{\mathbb{R}} \mathbf{1}\left(x | x \in r\right) x^2\mu_i(dx) - \nu_N^2 \right). \end{aligned} \tag{33}$$

A numerical example of the convolution $\lambda_N(r_N)$ is presented in Fig. 4. Theorems 2 and 3 can be directly generalized to $\mathbb{R}^n (n \in \mathbb{N})$, with $r \subset \mathbb{R}^d$.

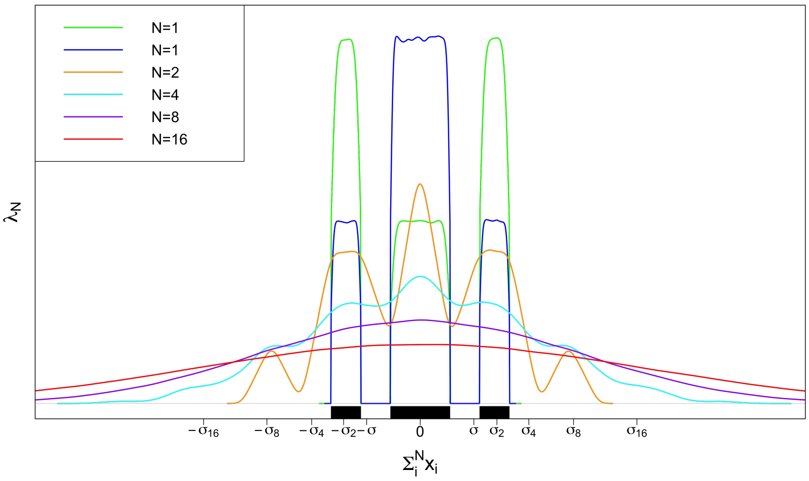

Figure 4. Numerical example of convolution $\lambda_N(r_N)$. For 2 kinds of probability measure μ_1 (green distribution) and μ_2 (blue distribution) on $r \subset \mathbb{R}$ (supported by black rug), the convolution $\lambda_N(r_N)$ with $N = 2, 4, 8, 16$ are shown with different colors based on random sampling of $600,000 \times N$ points from $\frac{N}{2}$ pairs of μ_1 and μ_2. The case of $\Lambda = 1$ is simulated, which shows the canonical convergence towards normal distribution following the central limit theorem with $\sigma_N \to \sqrt{N}\sigma$, where $\sigma = \frac{1}{2}(\beta_1 + \beta_2)$ as defined in (33) and (A13). For simplicity, ν_N is adjusted to 0 by the symmetric selection of μ_1, μ_2 and r.

2.3. Topological Structure of Complexity 1: Total Order of Observations

We consider the topological structure of inter-subjective objectivity based on the complexity defined as the convolution between different observations. As the commonality within inter-subjective objectivity is defined with multiple different observations, the topological ordering based on these complexity measures is possible with $N > 2$ observations of any nature.

We consider the commonality space with respect to each observation dataset as a point, and commonality between them as the distance between each pair of points. This can be considered as the undirected complete graph with N vertices, and its pair-wise complexity measure as ${}_NC_2$ edges length. The general property of Euclidean space allows a complete graph of size N to be embedded in $N-1$ dimensions (e.g., any line between 2 points is 1-dimensional space, and any triangle with 3 points is 2-dimensional surface, etc.), although an additional quantitative restriction such as triangle inequality on each triplet of edges is required. In order to treat an arbitrary set of the complexity measures and yield general characteristics of commonality space, we need to focus not on the actual values of complexity but on the topological order between them.

Let us first consider the total order between complexity values with $N > 2$ observation data contained in N vertices $V := \{v_i\}_{i=1,\cdots,N}$. One can determine the total order between ${}_NC_2$ edges $E := \{e_k\}_{k=1,\cdots,{}_NC_2} := \{v_i, v_{j\neq i} \in V\}$, by taking a mean order relationship between each pair of edges by the following algorithm (namely the pair-wise order algorithm):

1. For each pair of edges $\{e_i, e_{j\neq i} \in E\}$, calculate the order relation $e_i \leq e_j$ or $e_i \geq e_j$ with respect to the given complexity measure as an edge attribute such as length.
2. Score each edge e_i by mapping to integer $z : e_i \mapsto \mathbb{Z}$, by adding $+1$ if $e_i \geq e_{j\neq i}$ and by adding -1 if $e_i \leq e_{j\neq i}$, with respect to all other edges $e_{j\neq i}$.
3. The sorting with the score $\{z(e_i)\}$ provides the total order of E.

Note that the quantitative difference is completely lost in the case of antisymmetry, $(e_i = e_j) \equiv (e_i \leq e_j) \wedge (e_i \geq e_j)$. We will consider the meaning of this information loss with respect to other compatible sets of observation in the section 2.4.

Next, we consider the topological order of complexity for $N > 2$ observations according to the total order of these commonalities. We need here to translate the total order between edges E to that of observations V. This can be obtained by calculating the ${}_NC_3$ triplet of $N > 2$ vertices and associated total order of edges, with the following algorithm (namely triplet order algorithm schematicly represented in Fig.5):

1. For each triplet of observation $V_{i,j,k} := \{v_i, v_{j\neq i}, v_{k\neq i,j} \in V\}$ and associated edges $\{e_i := \{v_i, v_j\}, e_j := \{v_j, v_k\}, e_k := \{v_k, v_i\}\}$, update score of each observation by mapping to integer $z' : V_{i,j,k} \mapsto \mathbb{Z}$ with the following 6 rules:
2. If $e_i \geq e_j \geq e_k$ then $z'(v_i) = z'(v_i) - 1, z'(v_j) = z'(v_j) + 1, z'(v_k) = z'(v_k) + 0$.
3. If $e_i \geq e_k \geq e_j$ then $z'(v_i) = z'(v_i) + 1, z'(v_j) = z'(v_j) - 1, z'(v_k) = z'(v_k) + 0$.
4. If $e_j \geq e_i \geq e_k$ then $z'(v_i) = z'(v_i) + 0, z'(v_j) = z'(v_j) + 1, z'(v_k) = z'(v_k) - 1$.
5. If $e_j \geq e_k \geq e_i$ then $z'(v_i) = z'(v_i) + 0, z'(v_j) = z'(v_j) - 1, z'(v_k) = z'(v_k) + 1$.
6. If $e_k \geq e_i \geq e_j$ then $z'(v_i) = z'(v_i) + 1, z'(v_j) = z'(v_j) + 0, z'(v_k) = z'(v_k) - 1$.
7. If $e_k \geq e_j \geq e_i$ then $z'(v_i) = z'(v_i) - 1, z'(v_j) = z'(v_j) + 0, z'(v_k) = z'(v_k) + 1$.
8. The sorting with the score $\{z'(v_i) | i = 1, \cdots, N\}$ provides the total order of V.

The commonality order of V represents the topological structure of collective intelligence in citizen science with respect to inter-subjective objectivity, which corresponds to the topological inclusion relation of the Venn diagram in Fig. 1.

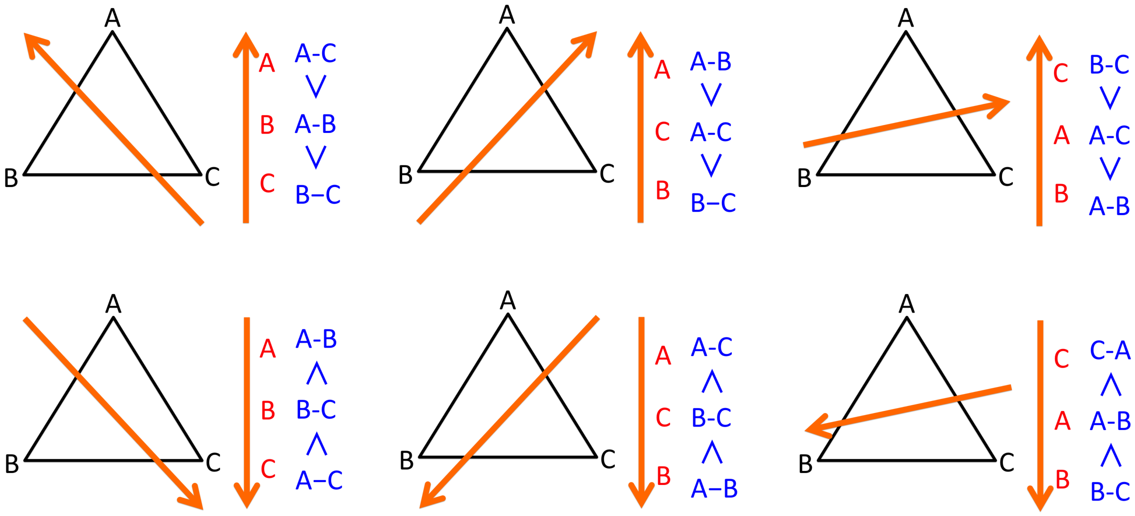

Figure 5. Schematic representation of the triplet order algorithm that calculates the total order of three observations with respect to the complexity defined on the pair-wise commonality between them. Three observations A, B and C are expressed as vertices of triangle in a 2-dimensional surface, whose edge lengths A-B, B-C, and A-C represent the commonality of each vertex pair. For simplicity, the triangles are projected as regular triangle, but the actual edge lengths generally differ, which provides the total order of edges. The six case statements of the algorithm are shown separately. Given the total order between the edges in blue magnitude relation, the corresponding total order of observations are depicted with orange axes at the side of each triangle. Orange axes superimposed with triangles signify that, by orthogonally projecting the vertices onto them, the total order of vertices are obtained, whose generalization is developed in the section 2.4. This holds for arbitrary three positive values of edge length without the constraint of triangular inequality, by considering appropriate projection of the triangles to a non-Euclidian surface.

2.4. Topological Structure of Complexity 2: Permutation between Total Orders of Observations

We expand the situation to 2 sets of $N > 2$ observations, namely observation I and II. For example, observer I and II observing N objects, or N observers observing 2 different objects I and II. It can also represent the application of two different complexity measures I and II to N observations. For simplicity, we limit the formalization to two sets of $N > 2$ observations, but generalization to a greater number of sets is possible.

In the general case, total orders I and II do not necessarily coincide. The relationship between 2 total orders with N observations can be described with the permutation of N elements (Fig.6(a)). In order to analyze the permutation between total orders, let $\mathcal{G}_N$ be a symmetric group with degrees of N. For $g \in \mathcal{G}_N$, we define a linear transformation $L_g : \mathbb{S}^N \mapsto \mathbb{S}^N$ by

$$L_g : (v_1, \cdots, v_N) \mapsto (v_{g(1)}, \cdots, v_{g(N)}), \tag{34}$$

which describes the permutation between commonality orders I and II.

We define a subspace $\mathbb{S}'(g)$ of $\mathbb{S}^N$ by

$$\mathbb{S}'(g) = \{v_i \in \mathbb{S} | v_i \neq v_{g(i)}\}, \tag{35}$$

which represents the subspace with compromise of total order. While by defining its complementary subspace

$$\mathbb{S}''(g) = \{v_i \in \mathbb{S} | v_i = v_{g(i)}\}, \tag{36}$$

we obtain the subspace in which there is no compromise, or the complete matching of two commonality orders. The whole commonality space can be divided into $\mathbb{S}'(g)$ and $\mathbb{S}''(g)$:

$$\mathbb{S}^N = \mathbb{S}'(g) \times \mathbb{S}''(g). \tag{37}$$

As depicted in Fig. 6(a) and (b), the compromise between 2 commonality orders is expressed as a non-linear folding relationship between them. Taking an assumption that the complexity measure is a continuous function, the integrated complexity measure that supports both commonality orders can be expressed as a folded structure, topologically speaking, such as the shape of the letter "N" (also the capital letter of Non-identical), taking the commonality measure of I and II as an Affine coordinate: An example with a red dotted line in Fig. 6(b) shows that we can compose an integrated commonality measure by bending the commonality measure II in an "N" shape with respect to that of I kept straight (in "I" shape, for Identical), which resolves the compromise. The "N" shape transformation of commonality measure means to change the topology of commonality order with respect to a permutation $g \in \mathcal{G}_N$ $(g(i) > g(j), 1 \leq i < j \leq N)$, while that of "I" shape represents the identical order $(g(i) < g(j), 1 \leq i < j \leq N)$. The non-compromising part of the two commonality orders conserves its order to the projection onto any linear combination of the two commonality measures, which topologically do not require "N" shape folding but maintain "I" shape matching.

For simplicity, We call the topological compromise between commonality orders the I-N compromise, and we call topologically identical matching I-I matching. Then I-I matching subspace $\mathbb{S}'(g)$ can be obtained as the linear combination of commonality measures I and II, and the subspace required for the resolution of I-N compromise corresponds to the complementary space $\mathbb{S}''(g)$ (Fig. 6(b) and (c)).

We call $\mathbb{S}'(g)$ an I-I space that consists of I-I dimensions, and $\mathbb{S}''(g)$ an I-N resolution space that consists of I-N resolution dimensions. The mean commonality order of two commonality orders projected onto I-I space (red solid arrows in Fig. 6(b) and (c)) can be obtained with the use of the pair-wise order algorithm in the section 2.3, applied not to commonality itself but to commonality orders. We call this the I-N mean commonality order, since it adopts the mean total order of commonality orders of I and II resolving the I-N compromise. Note that the information lost by antisymmetry of the pair-wise order algorithm does not affect the division of I-I and I-N resolution subspaces. Geometrical representation of the I-N compromise, I-I matching, and these corresponding dimensions, spaces and the I-N mean commonality order are given in Figs. 6.

We finally consider a statistical test on the degree of coincidence (TDC) between 2 commonality orders.

Theorem 4. *Statistical test on the degree of coincidence (TDC) between 2 commonality orders:*

Given that commonality orders I and II with N observations follow a uniformly random permutation with $\mathcal{G}_N$ as null hypothesis, the degree of coincidence d_c between the 2 commonality orders follows a binomial distribution:

$$\begin{aligned} k_{I\text{-}I} &:= \#\{(i,j) | g(i) < g(j), 1 \leq i < j \leq N, g \in \mathcal{G}_N\}, \\ P[d_c = k_{I\text{-}I}] &:= {}_M C_{k_{I\text{-}I}} p^{k_{I\text{-}I}} (1-p)^{N-k_{I\text{-}I}} \\ &\sim B(M,p), \end{aligned} \tag{38}$$

where $B(M,p)$ signifies binomial distribution of parameters $M = {}_N C_2$ and $p = 0.5$, $k_{I\text{-}I}$ represents the degree of coincidence as the number of I-I matching, $\#(\cdot)$ returns the size of the set, and $P[\cdot]$ the probability of the degree of coincidence d_c.

With respect to the **buoy-anchor-raft** model in section 1.2, the following correspondence is possible:

- 2 **observers observing N objects**: Commonality orders *I* and *II* can correspond to either of subjective (**buoy**) or objective (**anchor**) observation. The I-N resolution provides integrated commonality measure such as **buoy-anchor** connection and **raft** evaluation according to the nature of the observation. TDC provides connections between **buoy** and/or **anchors**.
- *N* **observers observing two different objects**: The commonality of *N* observers, whether it be subjective (**buoy**) or objective (**anchor**), are ranked with respect to two different objects *I* and *II*. The I-N resolution provides a mean ranking of *N* observers' commonality upon these observations. TDC provides the reproducibility of commonality among *N* observers.
- **Application of two different complexity measures to** *N* **observations**: For example, the case of **raft-anchor** connection where *N* subjective observers (**buoys**) are ranked with inter-subjective commonality (**raft** evaluation) and weighted with two different **anchors**. The I-N resolution provides mean ranking of *N* observers' inter-subjective objectivity integrating a multiple criteria of inter-subjective and objective evaluation. TDC represents statistical dependencies between 2 complexity measures in response to a given inter-subjective objective measurement. While significant matching between two commonality orders assures the reproducibility based on the coincidence of observation with these measures, non-significance can also be used to quantify complementarity of different evaluations [32].

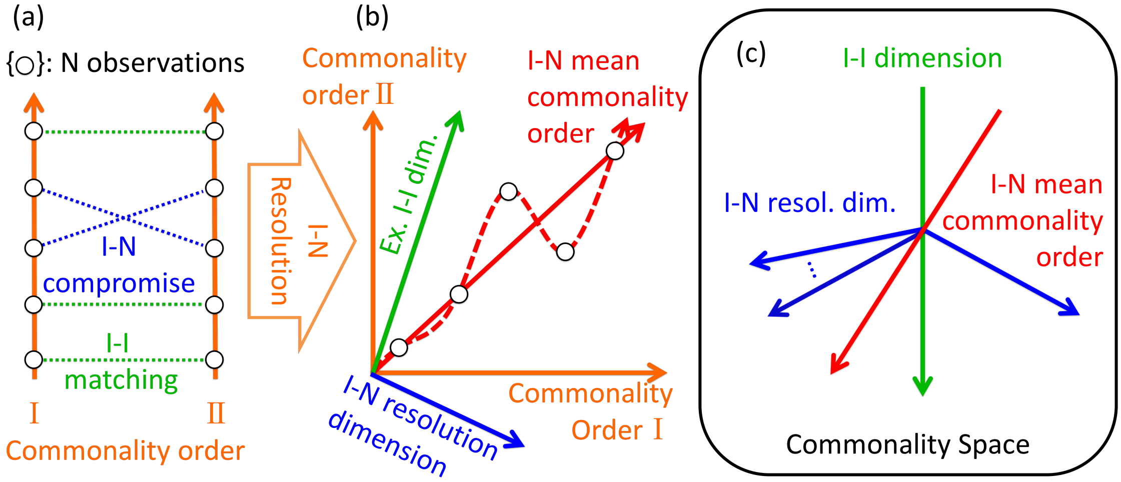

Figure 6. Integration of two commonality orders. (**a**): the correspondence between commonality orders *I* and *II* (orange arrows) can be described as the permutation between *N* observations (black circles), providing the topology of I-I matching (green dotted line) and I-N compromise (blue dotted line). (**b**): Affine space with respect to the commonality orders *I* and *II* as coordinate system (orange arrows) for the resolution of I-N compromise. The I-N mean commonality order (red solid arrow) can be calculated from the pair-wise order algorithm (section 2.3) applied on the commonality orders *I* and *II*, which makes the I-I matching identical to the I-I dimension (green arrow) and sets the mean order to I-N compromise. One I-N resolution dimension is required to resolve one I-N compromise (blue arrow). The implicit structure of the integrated commonality order with continuity assumption takes a complex form reflecting I-N compromises (red dotted arrow as an example), which corresponds to complex utility configuration in Fig.3(a). (**c**) The general case with an arbitrary number of I-N compromises. Total commonality space of $N-1$ dimensions is divided between I-N resolution dimensions (blue arrows) and I-I dimensions (green arrow), between which I-N mean commonality order can be defined (red arrow). $k < N$ axes of I-N resolution dimensions are required to resolve k I-N compromises (blue arrows). Taking the I-I dimensions and I-N resolution dimensions as Affine coordinates, the integrated commonality order is projected onto the I-N mean commonality order as a simplest sorted order of utility, which corresponds to Fig. 3(b).

3. Computational Complexity

The computation of complexity measures and commonality orders depends on exhaustive calculation of combinatorics between observations. The computational complexity of such calculation should also be investigated in terms of topological complexity, in order to yield a general theoretical platform that does not depend on the particularity of the database.

3.1. Topological Complexity of Commonality

First, we investigate topological order of commonality among N observations. Using the convolution as commonality (27), we define the maximum commonality order $O : \mathbb{X} \mapsto \mathbb{N}$ as follows:

$$O(r \subset \mathbb{X}) := max\{k \in 1, \cdots, N | \lambda_k(r) > 0\}. \tag{39}$$

The general topological structure of $O(\mathbb{X})$ is depicted in Fig. 7.

On the cardinality of $O(\mathbb{X})$, the following holds:

Theorem 5. *As $\#(\mathbb{X}) \to \aleph_0$, $^\exists r \subset \mathbb{X}$ such that $\#(\{r | O(r) = \infty\}) = \aleph_0$, where $\aleph_0$ represents aleph-naught.*

This means that for any elaborated inter-subjective objectivity, there is always the possibility to develop another different set of observations that attains higher inter-subjective objectivity by increasing the dataset. This structure assures the representation of a paradigm shift in science when sufficient contradicting evidence gained a majority compared to an old model. For example, minority reports in biology that may lead to novel discoveries in the future can be properly stored and distinguished from erroneous reports as more evidence accumulates [27].

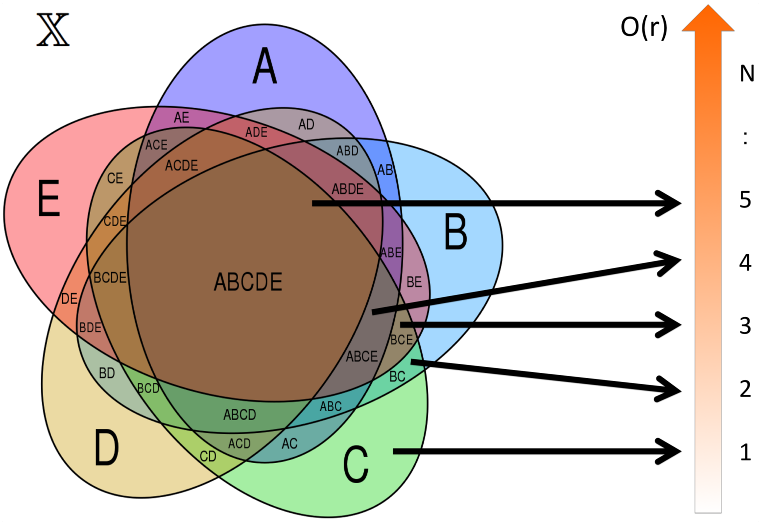

Figure 7. Topological hierarchy of commonality between observations. For example, 5 observations A, B, C, D, E are depicted with correspondence to commonality order of each topological subset. The Venn diagram on the left represents the commonality structure within observation probability database $\mathbb{X}_5$ on variable $\mathbb{X}$ ($N = 5$ in section 3.2), where coincident observation is superimposed. The maximum commonality order is the projection between these topological subsets to the natural number $\mathbb{N}$ in right axis, describing the number of matching observations. Venn diagram cited from [34].

3.2. Algorithmic Complexity

Secondly, we evaluate the computational complexity with respect to the computing time scale. Since data-driven citizen science requires real-time computation in a highly interactive manner with observation process, the algorithmic complexity of the calculation of complexity measures is an essential limiting factor of performance. As commonality is based on the intersection of multiple observations, its exhaustive computing confronts combinatorial explosion as datasets increase. Although computation of complexity itself, or resolution of search query as mathematical theorem is provable and an algorithmic solution can be found, the computation resource is another practical issue for real world implementation, especially in distributed observation.

The computational time scale required for the sorting of a database according to a given utility such as commonality is listed in Table 3. Under a general condition with the observation probability database $\mathbb{X}_N$ of size N, $\mathbb{X}_N := \{\mu_i(x) | x \in \mathbb{X}, i = 1, \cdots, N\}$, maximum complexity lies in the calculation of commonality order based on the intersection of $\left\lfloor \frac{N}{2} \right\rfloor$ or $\left\lceil \frac{N}{2} \right\rceil$ elements, whose sorting time belongs to factorial order of N . The case with $N = 5$ is depicted in Fig. 7. This means that an algorithmic burden exists towards the calculation of middle-scale commonality with respect to the data size. As an inter-subjective objectivity successfully increases in citizen science, this peaking of algorithmic complexity in intermediate scale may hinder the effective feedback necessary for guided self-organization.

However, in a practical situation, the actual computation time may remain in polynomial order if effective data size shrinks with respect to the increase of maximum commonality order:

Theorem 6. *By defining the diminution rate of data combination $\Delta : \mathbb{N} \mapsto \mathbb{R}$ with respect to maximum commonality order $1 \leq i \leq N \in \mathbb{N}$ as*

$$\Delta(i) := \frac{{}_N C_{\dim(\{\mathbb{X}_N | O(\mathbb{X}) = i\})}}{{}_N C_{\dim(\{\mathbb{X}_N | O(\mathbb{X}_N) = i-1\})}} \tag{40}$$

the order of its product is upper bounded by the d-th root of maximum computational complexity at $N' = \left\lfloor \frac{N}{2} \right\rfloor$

$$\prod_{i}^{N'} \Delta(i) \leq \sqrt[d]{\mathcal{O}(N^{dN'})}, \tag{41}$$

where $\dim(\cdot)$ returns the size of the database, and $d > 0$ represents the polynomial order of the algorithm $\mathcal{O}(N^d)$ with respect to the data size N.

From this result, we can conjecture that for $N'' \leq N'$,

$$\prod_{i}^{N''} \Delta(i) \leq \mathcal{O}(N^{\frac{c}{d}}) \tag{42}$$

will assure exhaustive feedback with polynomial response time of degree c. Usually, the left side is based on the past calculation of lower maximum commonality order, we can annotate interactively whether interactive information processing can assure comprehensive feedback. This will add a criterion on the criticality of guided self-organization mediated by computation, which will be explored in the section 4.

Another methodology other than exhaustive computing is to implement local gradient algorithm as local interaction that leads to global heuristic solution without top-down control. This can also be achieved with the use of limited maximum commonality order, for example, $O(\mathbb{X}) = k < N'$, which will keep its computational time within polynomial order $\mathcal{O}(N^{dk})$.

Table 3. Algorithmic complexity for the calculation of commonality orders. With respect to the maximum commonality order in (39), an exhaustive number of combinations with the use of observation probability database $\mathbb{X}_N$ of size N and the time scale required for the sorting of the commonality measure is shown. Sorting time is based on the worst case performance of canonical algorithms such as bubble sort and quick sort (polynomial degree $d = 2$). $\mathcal{O}(\cdot)$ denotes asymptotic notation of Landau. $O(\mathbb{X}) = \left\lfloor \frac{N}{2} \right\rfloor$ and $\left\lceil \frac{N}{2} \right\rceil$ require the maximum calculation and sorting time. Note that the total computation time is upper bounded by the sorting process ($d = 2$) than the combinatorics of commonality ($d = 1$), though calculation time of each commonality such as convolution should be further considered in actual implementation.

Maximum Commonality Order $O(\mathbb{X})$	**Number of Combination**	**Sorting Time (**$d = 2$**)**
2	${}_NC_2$	$\mathcal{O}(({}_NC_2)^2) = \mathcal{O}(N^4)$
3	${}_NC_3$	$\mathcal{O}(({}_NC_3)^2) = \mathcal{O}(N^6)$
$\vdots$	$\vdots$	$\vdots$
$\left\lfloor \frac{N}{2} \right\rfloor$ or $\left\lceil \frac{N}{2} \right\rceil$	${}_NC_{\lfloor \frac{N}{2} \rfloor} = {}_NC_{\lceil \frac{N}{2} \rceil}$	$\mathcal{O}\left(\left({}_NC_{\lfloor \frac{N}{2} \rfloor}\right)^2\right) = \mathcal{O}\left(\left({}_NC_{\lceil \frac{N}{2} \rceil}\right)^2\right) = \mathcal{O}\left(N^{2\cdot\lfloor \frac{N}{2} \rfloor}\right)$
$\vdots$	$\vdots$	$\vdots$
N	${}_NC_N = 1$	$\mathcal{O}(({}_NC_N)^2) = \mathcal{O}(1)$

3.3. Big Data Integration

Thirdly, we consider the computational complexity required for big data integration. As open data is increasingly gaining its availability in citizen science, integration of massive databases from different resources has become one of the most important data processing methods. The conversion of different databases through application programming interface is a basic protocol when the database is distributed over multiple servers.

The computation required in big data integration is the extensive calculation of commonality in the direct product of multiple databases. For simplicity, we consider the integration of 2 databases $\mathbb{X}_N$ and $\mathbb{X}_M$, with size N and $M \in \mathbb{N}$, $\mathbb{X}_N := \{\mu_i(x) | x \in \mathbb{X}, i = 1, \cdots, N\}$, $\mathbb{X}_M := \{\mu_i(x) | x \in \mathbb{X}, i = 1, \cdots, M\}$, respectively. A joint distribution between subsets of $\mathbb{X}_N$ and $\mathbb{X}_M$ needs to be determined with respect to common parameters, in order to obtain integrated database including the calculation of up to $(N + M)$-th order of commonality, such as order-wise correlations [32]. Exhaustive computing follows the argument in section 3.2, giving the extension of the theorem 6:

Theorem 7. *Given the diminution rate of data combination* $\Delta' : \mathbb{N}^2 \mapsto \mathbb{R}$, *with respect to maximum commonality order* $1 \leq i \leq N \in \mathbb{N}$ *and* $1 \leq j \leq M \in \mathbb{N}$, *during integration of* 2 *databases* $\mathbb{X}_N$ *and* $\mathbb{X}_M$, *respectively, as*

$$\Delta'(i,j) := \frac{{}_NC_{\dim(\{\mathbb{X}_N | O(\mathbb{X}) = i\})}}{{}_NC_{\dim(\{\mathbb{X}_N | O(\mathbb{X}) = i-1\})}} \cdot \frac{{}_NC_{\dim(\{\mathbb{X}_M | O(\mathbb{X}) = j\})}}{{}_NC_{\dim(\{\mathbb{X}_M | O(\mathbb{X}) = j-1\})}}, \tag{43}$$

the order of its product is upper bounded by the d-th root of maximum computational complexity at $N' = \left\lfloor \frac{N}{2} \right\rfloor$ *and* $M' = \left\lfloor \frac{M}{2} \right\rfloor$

$$\prod_{(i,j)}^{(N',M')} \Delta'(i,j) \leq \sqrt[d]{\mathcal{O}([N^{N'} M^{M'}]^d)}, \tag{44}$$

where $d > 0$ *represents the polynomial order of the algorithm* $\mathcal{O}([N^{N'} M^{M'}]^d)$ *with respect to the data size* N *and* M.

In this formalization, computational complexity of database integration also confronts combinatorial explosion with respect to data size. Similarly to (42), we then explore a practical condition that effective maximum commonality order can be treated with polynomial time of degree $c > 0$, such that

$$\mathcal{O}([N^{N'}M^{M'}]^{d}) \leq \mathcal{O}([N+M]^{c}). \tag{45}$$

For that purpose, we set the uniform sparseness u $(0 < u < 1)$ of random databases representing the density of combination that supports the existence of commonality at each order,

$$\begin{aligned} \frac{{}_{N}C_{\dim(\{\mathbb{X}_N|O(\mathbb{X})=k\})}}{{}_{N}C_{k}} &= \frac{{}_{M}C_{\dim(\{\mathbb{X}_M|O(\mathbb{X})=k\})}}{{}_{M}C_{k}} = u \\ \text{for } k &= 1, \cdots, N' \text{ or } M', \end{aligned} \tag{46}$$

which maintains the diminution rate of data combination Δ (40) and Δ' (43) invariant under the definition. With respect to the total size of the database after integration $L = N + M$, the following holds:

Theorem 8. *As $L \to \infty$ in random data (46), the mean condition of (45) for all $\{N, M | N + M = L\}$ converges to the following inequality, which represents polynomial time constraints on computational complexity for exhaustive calculation of newly emerging commonality order within data size L:*

$$u \leq \mathcal{O}(f * f(L)), \tag{47}$$

where

$$f(x) := \frac{L^{\frac{c}{4d}} x^{-\frac{L}{8}}}{\sqrt{L}}, \tag{48}$$

and $*$ signifies the discrete convolution (20):

$$f * f(L) := \sum_{N=1}^{L-1} f(N) f(L-N). \tag{49}$$

Numerical observation of the proof is given in Fig. 8.

This signifies that the convolution of power function of each database's size serves as the complexity measure of big data integration with respect to computational complexity. This provides the condition of data sparseness u such that exhaustive calculation of all newly generating commonality orders within size L can be treated with polynomial time order c under algorithmic constraint d. As the inequality indicates, the more data is sparse, the easier we can calculate joint commonality.

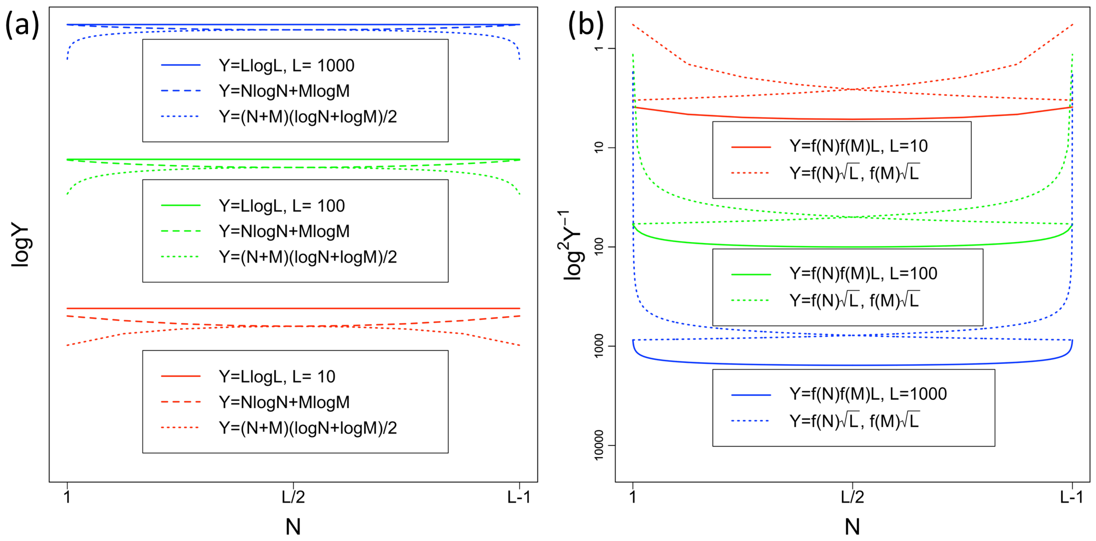

Figure 8. Numerical observation of the proof of theorem 8. (**a**) Chebyshev's inequality (A41) and asymptotic convergence to $\mathcal{O}(L \log L)$ (A44) with respect to $N, M \geq 1$ $(N + M = L)$, $L = 10, 10^2, 10^3$. Y-axis is plotted with log scale. The equality in (A41) is given at $N = M = \frac{L}{2}$. (**b**) Behaviour of $f(N)\sqrt{L}$, $f(M)\sqrt{L}$ and $f(N)f(M)L$ with respect to $L = 10, 10^2, 10^3$. For visibility, Y-axis scale is given as $\log^2(Y^{-1})$ that represents smaller Y value to the bottom, and Y-axis label shows the value of $-\log Y$. The surface below the solid line $f(N)f(M)L$ represents the convolution multiplied by L, $f * f(L)L$. The mean value of solid line $f(N)f(M)L$ therefore corresponds to the upper limit of u that satisfies the polynomial constraint (45) with respect to given L. $c = d = 2$ were used for the simulation.

4. Conjectures on Guided Self-Organization

With effective feedbacks by computation, citizen science dynamics is expected to converge to a critical state where objective is collectively optimized through the mutual increase of inter-subjective objectivity. However, several aspects may intervene in the resulting self-organized state, on which we need theoretical interpretation. In this section, general important aspects are exemplified in relation to self-organized criticality.

4.1. Criticality by Limitation

The accuracy and reproducibility of observation is a primary factor that defines the consequent resolution of information represented in a database. Computational complexity also gives constraint on the speed of information processing for prediction. Several limiting factors may generically arise, such as:

1. **Limitation by principle**: Deterministic chaos inherent in a natural system does not allow for long-term prediction, because the tiniest observation error of present state will develop in exponential order [35]. Short-term validity of meteorologcal prediction is a typical example.
2. **Limitation by reproducibility**: In a real world situation, we mostly encounter one-time-only events, which do not allow reproduction under the same condition [5]. Available data is sparse with respect to latent variables, which causes quantitative limitation of prediction [4].
3. **Limitation by computational complexity**: As explored in section 3.2, extensive feedback based on exhaustive computing is often impossible with respect to available computing resources. The resolution of feedback may include time delay or incomplete optimization. Spatial-temporal scale of the forecast also sets the constraint as a general trade-off between prediction accuracy and computational resources. The coarser the forecast granularity is, the more costly the calculation becomes but the more likely it is to realize an accurate long term prediction.

These limitations fundamentally regulate the order of significant digits in the prediction process, at the edge of resulting precision where the accuracy reaches criticality. The whole dynamics is also confined by the criticality of the observing phenomena itself, by which observers' behaviour is influenced.

4.2. Criticality by Successful Learning

The motivation of citizen science is not necessarily the construction of versatile artificial intelligence, but the integration and augmentation of human capacity as well [4] [13] [12]. Successfulness of citizen science can also be defined in terms of information transition from machine to human, on which criticality is assumed to appear.

Let us consider the case when successful learning mediated by computation transferred effective prediction model into human cognitive capacity. We take an example with Bayesian estimation, which is also a general model of our brain function [36]. General formulation of Bayesian estimation updates the parameter of hypothesized prior probability $P(A)$ with respect to the observed data $P(B)$, and provide estimation of posterior probability $P(A'|B)$ given by Bayes' theorem:

$$P(A'|B) := \frac{P(B|A)P(A)}{P(B)}, \tag{50}$$

where $P(B|A)$ is considered as likelihood function, which updates $P(A)$ to $P(A'|B)$.

We now consider that the prior probability $P(A)$, or the model of prediction, depends on the process of computation C and human decision D. As human decision is supported by computation,

$$P(A) := P(D|C). \tag{51}$$

This formalization corresponds to Bayesian hierarchical modelling, where computation C provides hyperparameter of human decision D as prior distribution:

$$\begin{aligned} P(D,C|B) &:= \frac{P(B|D)P(D,C)}{P(B)}, \\ &:= \frac{P(B|D)P(D|C)P(C)}{P(B)}. \end{aligned} \tag{52}$$

When human successfully acquired the model represented in computational model,

$$P(D|C) \approx P(D) \tag{53}$$

as independent identical distribution, and

$$P(D) \sim P(C) \tag{54}$$

as independent and informationally homologous distribution.

This criticality qualitatively corresponds to the saturation stage of Markov chain Monte Carlo method (MCMC) in the optimization of hierarchical model (52), where hyperparameter and parameter converge to independent stable distributions. Therefore, by monitoring the dependency of machine-human interaction with respect to the actual predictability, one can suggest whether the computation model or human observation should change, or if the actual phenomenon is in transition:

- **When the actual prediction accuracy is high and human-machine interaction is high**, this indicates the successful modelling of observing phenomenon with the use of computation.
- **When actual prediction accuracy is high and human-machine interaction is low**, it means the human has achieved a successful understanding of the phenomenon with less dependency on a machine.
- **When actual prediction accuracy is low and human-machine interaction is high**, it indicates the possibility that computational capacity is not sufficient to effectively treat the phenomenon. Otherwise, the observing phenomenon might be in dynamical transition that effective computational model needs to be changed.
- **When actual prediction accuracy is low and human-machine interaction is low**, more human effort needs to be engaged both on actual observation and the utilization of machine interface.

4.3. Criticality by Guided Optimization

Actual management task of citizen science is often firmly related to the sustainability of social-ecological system, where achievement of robustness and resilience is an important criterion of criticality [3] [5]. Universally robust model with respect to arbitrary variable cost function is canonically given by uniform distribution, which is commonly adopted as a prior of Bayesian estimation and random search algorithm [14]. It is also widely prevalent in biological phenomena as the survival rate depends on the geometric mean of evolutionary fitness, which is maximized with uniformity in space, time and statistical configuration [32] [37].

On the other hand, a short-term management goal is usually biased by a given objective. How to reconcile short-term local efficiency and long-term global sustainability is a crucial issue for guided self-organization of management in citizen science.

In order to optimize the balance between different spatio-temporal scales, information geometry can provide theoretical compromise in terms of informational complexity. Suppose actual distribution of variable $X \subset \mathbb{X}$ is given by $P_a(X)$, a short-term management goal as $P_s(X)$ and idealized long-term robust distribution as $P_l(X)$. In many natural systems, the uniformity of $P_l(X)$ supporting robustness as the result of self-organization is expressed with entropy maximization principle under parameter constraints such as resource availability and energy flux level [38].

For simplicity, take an example with Shannon's diversity index H' defined on discrete distribution $P(X)$ on symbols $X = \{s_0, s_1, \cdots, s_n\}$, such as frequency of n species in biodiversity observation.

$$H' := -\sum_{i=0}^{n} P(s_i) \log P(s_i), \tag{55}$$

where s_0 represents the non-occurrence of any species. $P(X)$ and H' could be either **buoy** or **anchor**. Note that H' can be generalized to mutual information H'_2 to express **raft**, **buoy-anchor** and **raft-anchor** connections,

$$H'_2 := \sum_{i,j} P_2(s_i, s_j) \log \frac{P_2(s_i, s_j)}{P(s_i)P(s_j)}, \tag{56}$$

where $P_2(\cdot, \cdot)$ denotes joint distribution on $X \times X$.

By maximizing H', we can determine the most diverse distribution P_l as

$$P_l(s_i) = \frac{1}{n+1}, \tag{57}$$

which represents the most robust ecosystem taking on the assumption that every species including the gap is equally invaluable in terms of ecosystem function in randomly changing environment.

With respect to the subset of X we focus for short-term management goal, both $H'(P_a) < H'(P_s)$ and $H'(P_a) > H'(P_s)$ could occur. However, a general relationship between biodiversity and ecosystem functions imposes $H'(P_a) < H'(P_s)$, meaning a net positive impact on biodiversity and good management in terms of sustainability. H' can be generalized to complexity measure G^{-1} in section 2.1 with respect to the commonality λ in section 2.2, which will be detailed in section 6.

Expressed as an exponential family, $P(X)$ can be parameterized as a statistical manifold based on the canonical setting of information geometry, with the dual-flat coordinates $\Theta = \{\theta_i | i = 1, \cdots, n\}$ and $H = \{\eta_i | i = 1, \cdots, n\}$, with potential functions ϕ and ψ, respectively, based on the Fisher information metric g and connection coefficients $\Gamma^{(\alpha)}$ [39][40]:

$$\begin{aligned} P(X, \Theta) &= \exp\left[C(X) + \sum_{i=1}^{n} \theta_i f_i(X) - \psi(\Theta)\right], \\ \frac{\partial}{\partial \theta_i}\psi &= \eta_i, \\ \frac{\partial}{\partial \eta_i}\phi &= \theta_i, \\ \phi(H) &= \sum_{i=1}^{n} \theta_i \eta_i - \psi(\Theta), \end{aligned} \tag{58}$$

under the correspondence of the following transformation for discrete distribution,

$$\begin{aligned} C(X) &= 0, \\ f_i(X) &= \mathbf{1}(X | X = s_i), \\ \psi(\Theta) &= -\log P(s_0), \\ \theta_i &= \log \frac{P(s_i)}{P(s_0)}, \\ \eta_i &= E[f_i(X)] = P(s_i). \end{aligned} \tag{59}$$

The elements of Fisher information metric $g = (g_{ij})$ are given with respect to the dual coordinates,

$$\begin{aligned} g_{ij} &= \frac{\partial}{\partial \theta_i}\frac{\partial}{\partial \theta_j}\psi(\Theta) &= \frac{\partial \eta_j}{\partial \theta_i}, \\ g^{inv}_{ij} &= \frac{\partial}{\partial \eta_i}\frac{\partial}{\partial \eta_j}\phi(H) &= \frac{\partial \theta_j}{\partial \eta_i}, \end{aligned} \tag{60}$$

where $\left(g^{inv}_{ij}\right)$ is the inverse matrix of (g_{ij}). This relation defines Θ and H as the dual coordinate systems orthogonal to each other with respect to g. The α-connection coefficients $\Gamma^{(\alpha)} = \left(\Gamma^{(\alpha)}_{ij;k}\right)$ $(i, j, k \in \{1, \cdots, n\})$ with respect to a real number α is given by the Fisher information metric as

$$\Gamma^{(\alpha)}_{ij;k} = \frac{1}{2}\left(\frac{\partial}{\partial \theta_i} g_{jk} + \frac{\partial}{\partial \theta_j} g_{ik} + \frac{\partial}{\partial \theta_k} g_{ij} - \alpha E\left[\frac{\partial}{\partial \theta_i}\log P(X)\frac{\partial}{\partial \theta_j}\log P(X)\frac{\partial}{\partial \theta_k}\log P(X)\right]\right), \tag{61}$$

where $E[\cdot]$ is the mean value function. The values $\alpha = 1$ and -1 are essential in information geometry, which define the e- and m-flat connections respectively, in terms of the invariance of tangent space under the covariant differential $\nabla^{(\alpha)}$ on arbitrary coordinates $\{\xi_i\}(i = 1, \cdots, n)$ of the statistical manifold:

$$\nabla^{(\alpha)}_{\frac{\partial}{\partial \xi_i}} \frac{\partial}{\partial \xi_j} = \sum_{k=1}^{n} \Gamma^{(\alpha)}_{ij;k} \frac{\partial}{\partial \xi_k}, \tag{62}$$

where $\Gamma^{(1)}_{ij;k} = 0$ for $\xi_i = \theta_i$, and $\Gamma^{(-1)}_{ij;k} = 0$ for $\xi_i = \eta_i$. For example, the model $P(X;\Theta)$ is e-flat with respect to the coordinates Θ, and m-flat with respect to the coordinates H. $\nabla^{(\pm 1)}$ is called the dual-flat connection of the statistical manifold. The concept of flatness defined by these connections further extends to the concept of geometric parallel and geodesic. As an autoparallel submanifold with respect to the connection, e- and m-flat geodesic $\Theta(w)$ and $H(w)$ between 2 distributions $P_1(X)$ and $P_2(X)$ are defined as follows with one-dimensional parameter w:

$$\Theta(w) = w\Theta(P_1(X)) + (1-w)\Theta(P_2(X)), \tag{63}$$

$$H(w) = wH(P_1(X)) + (1-w)H(P_2(X)). \tag{64}$$

The unique $\nabla^{(\alpha)}$-divergence $D^{(\alpha)}(P_1(X) : P_2(X))$ that satisfies $D(P_1(X) : P_2(X)) \geq 0$ and $D(P_1(X) : P_2(X)) = 0 \Leftrightarrow P_1(X) = P_2(X)$, and that remains invariant under possible transformations of the dual-flat coordinates with the connections $\nabla^{(\pm\alpha)}$ is given by

$$D^{(\alpha)}(P_1(X) : P_2(X)) = \Psi(P_1(X)) + \Phi(P_2(X)) - \sum_{i=1}^{n} \theta_i(P_1(X))\eta_i(P_2(X)), \tag{65}$$

whose dual divergence coincides with Kullbuck-Leibler divergence in case of $\alpha = 1$,

$$D^{(1)}(P_1(X) : P_2(X)) = D^{(-1)}(P_2(X) : P_1(X)) = \sum_X P_2(X) \log \frac{P_1(X)}{P_2(X)}. \tag{66}$$

From the Pythagorean relation and the projection theorem of Kullbuck-Leibler divergence on the dual-flat statistical manifold [39](p.63), the following holds:

Theorem 9. *Let $(\Theta_a, H_a), (\Theta_s, H_s)$ and (Θ_l, H_l) be the dual-flat coordinates of $P_a(X), P_s(X)$ and $P_l(X)$, respectively, with canonical definition of e- and m-flat dual connections. We define the optimal distribution $P_o(X)$ with coordinates (Θ_o, H_o) on m-flat geodesic between $P_a(X)$ and $P_l(X)$ with parameter $w \in \mathbb{R}$ as*

$$H_o \quad := \quad wH_a + (1-w)H_l. \tag{67}$$

By optimizing H_o with orthogonal projection of e-flat geodesic from Θ_s to Θ_o as

$$w = \arg\min_w (D^m(P_o : P_s)) = \arg\min_w (D^e(P_s : P_o)), \tag{68}$$

the following Pythagorean relations hold:

$$\begin{aligned} D^m(P_a : P_s) &= D^m(P_a : P_o) + D^m(P_o : P_s), \\ D^m(P_l : P_s) &= D^m(P_l : P_o) + D^m(P_o : P_s). \end{aligned} \tag{69}$$

Where $D^m(\cdot : \cdot)$ and $D^e(\cdot : \cdot)$ are Kullback-Leibler divergence and its dual divergence, respectively,

$$D^m(P_o : P_s) := D^e(P_s : P_o) := \sum_{i=0}^{n} P_o(s_i) \log \frac{P_o(s_i)}{P_s(s_i)}. \tag{70}$$

Fig. 9 shows the geometrical structure of this theorem. In this case, supposing $H'(P_a) < H'(P_s) < H'(P_l)$ as effectiveness of complexity measure H' for management, we want to find optimal distribution of biodiversity P_o balancing between P_s and P_l with respect to actual distribution P_a, such that

$$H'(P_a) < H'(P_s) < H'(P_o) < H'(P_l), \tag{71}$$

based on statistical dependencies between variables that can be orthogonally separated with Pythagorean relation. As a result, P_o provides the optimized distribution with respect to minimum informational discrepancy from the short-term objective to the ideal transition towards the long-term most diverse state. The meaning of major components of Kullbuck-Leibler divergence to be used as a guide of self-organization is listed as follows:

- $D^m(P_a : P_o)$: Discrepancy between actual distribution and optimum portfolio strategy that orthogonally decomposes and attempts to achieve a balance between short-term management objective and long-term sustainability.
- $D^m(P_a : P_s)$: Target risk of short-term management objective.
- $D^m(P_o : P_s) = D^e(P_s : P_o)$: Buffering element of robustness trade-off between short-term management objective and long-term sustainability.
- $D^m(P_l : P_o)$: Potential risk of optimum portfolio w.r.t. long-term sustainability.
- $D^m(P_l : P_s)$: Potential risk of short-term management objective w.r.t. long-term sustainability.
- $D^m(P_l : P_a)$, $D^m(P_a : P_l)$: Potential risk of actual distribution w.r.t. long-term sustainability.

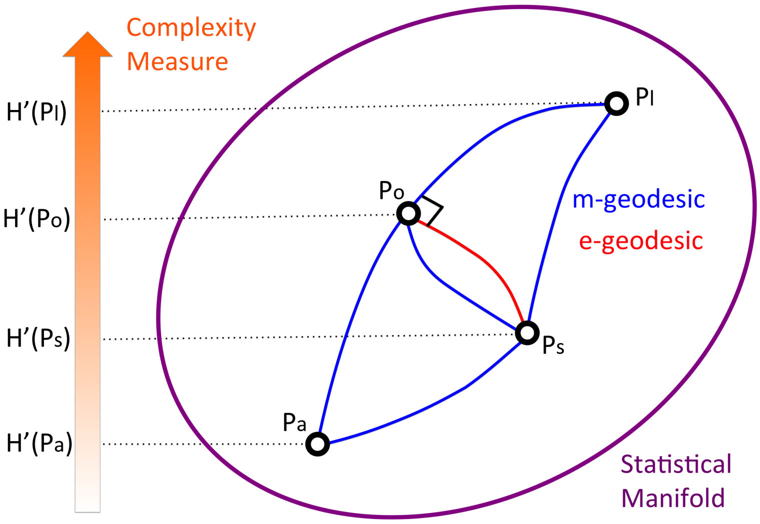

Figure 9. Information geometrical optimization of diversity strategy portfolio with respect to actual distribution P_a, short-term management objective P_s and long-term sustainability P_l. On a dual-flat statistical manifold based on Fisher information metric, each distribution is represented as a point (black circles). The m-geodesic is depicted with blue line, while e-geodesic is shown with a red line, which orthogonally cross at the optimized strategy P_o. Topological correspondence between complexity measure H' (aligned on left orange arrow) and diversity strategy portfolio (P_a, P_s, P_l and P_o) is shown with dotted lines with respect to the magnitude relation.

5. Results from Biodiversity Management

We demonstrate the application of the model developed in this article to actual citizen science observation data, taking a biodiversity observation activity supported by interactive database as a typical example [17]. Sample data contain the observation by 7 citizen participants on 48 subjective binary indices on species occurrence as **buoy** data on biological diversity, resulting in 336 samples. On the other hand, a **buoy-anchor** connection was established separately by objective evaluation of each participant's ability to detect these species.

Commonality orders among seven observers were obtained for both inter-subjectivity based on the mutual information of **buoy** data, and subjective-objective unity by simply ranking with **buoy-anchor** connection data. These orders are shown in Fig. 10. A binomial test defined in (38) was performed on the comparison between inter-subjective and subjective-objective commonality orders. The random order distribution hypothesis was rejected with respect to 4% significance threshold. The matching was more consistent in a higher order of commonality, which implies the intervention of subjective bias in a lower order. With respect to the conjectures on criticality in section 4, the results can be interpreted as a significant self-organization process towards criticality with the increase of inter-subjective objectivity.

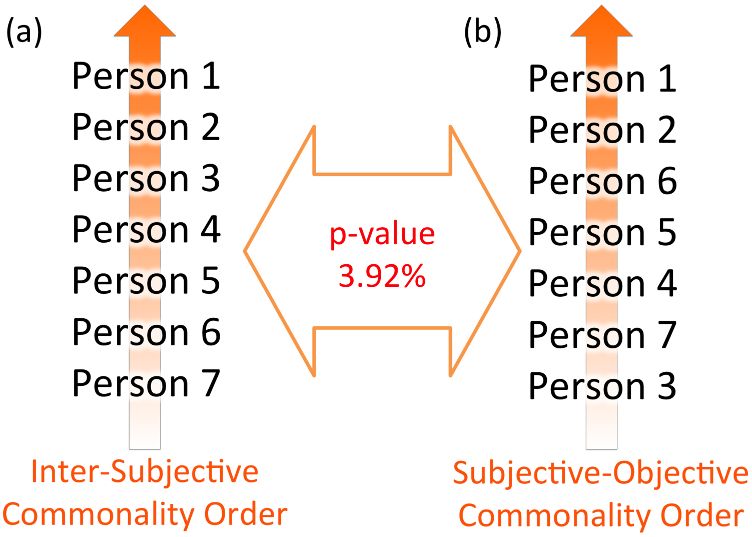

Figure 10. Results of inter-subjective and subjective-objective commonality orders in citizen observation of biodiversity. Seven people represented with numerical ID are aligned with commonality orders, (**a**): based on inter-subjectivity, and (**b**): based on subjective-objective unity, which showed a 3.92% residual error probability regarding the rejection of the random order distribution hypothesis with respect to the binomial test (38).

6. Discussion

We have tackled the general situation in data-driven citizen science where scientific accuracy and reproducibility can only be discussed at the intersection of subjectivity, inter-subjectivity, and objectivity. Based on the conceptual definition of inter-subjective objectivity, a general topological structure was characterized with respect to complexity measure, search function, computational complexity and criticality conditions. The results provide theoretical criteria for the development of information and communication technology in view of effective assistance and guidance of citizen science from a complex systems perspective.

The universality of the developed theory and models lies in the generality of the commonality concept formalized as convolution. In reality, a joint distribution of N variables can be represented as the function of convolution with degree N, which allows for extensive expression of informational complexities [32].

For example, by choosing the time range $T \subset \mathbb{R}$ with positive Lebesgue measure $m(T) > 0$, marginal distribution $P(x|T)$ can be expressed as the time integral of probability measure μ, such as

$$\begin{aligned} P(x|T) &:= \int_T \mu(dt) \\ &= \int_{\mathbb{R}} \mathbf{1}(t|t \in T)\mu(dt) \\ &= q(T), \end{aligned} \tag{72}$$

according to (24).

On the other hand, joint distribution $P(x_1, x_2|T)$ is also the time integral of the products between each variable's probability measure μ_1 and μ_2, within simultaneous time range $\{dT^i\}_{i=1,\cdots,n}$:

$$\begin{aligned} P(x_1, x_2|T) &:= \sum_i^n \int_{dT^i} \int_{dT^i} \mu_1(dt_1)\mu_2(dt_2)m(dT^i), \\ \bigcup_i^n dT^i &= T. \end{aligned} \tag{73}$$

where $m(\cdot)$ is Lebesgue measure on $\mathbb{R}$. As defined in (25),

$$\begin{aligned} \int_{dT^i} \int_{dT^i} \mu_1(dt_1)\mu_2(dt_2) &= \mu_1 * \mu_2(dT_2^i), \\ dT_2^i &:= \left\{ \sum_{i=1,2} t_i \middle| t_i \in dT^i \right\}, \end{aligned} \tag{74}$$

which derives the practical form for actual data processing as

$$P(x_1, x_2|T) := \sum_i^n \mu_1 * \mu_2(dT_2^i)m(dT^i). \tag{75}$$

Taking $n \to \infty$, we obtain

$$\begin{aligned} P(x_1, x_2|T) &:= \int_T \mu_1 * \mu_2(dT_2) \\ &= \int_T q_1(dT)q_2(dT) \\ &= \int_T \mu_1(dT)\mu_2(dT), \end{aligned} \tag{76}$$

the canonical definition of joint distribution with real value resolution of time.

This follows the generalization to N variables with (A5) as

$$\begin{aligned} P(x_1, x_2, \cdots, x_N|T) &= \sum_i^n \lambda_N(dT_N^i)m(dT^i), \\ dT_N^i &:= \left\{ \Lambda \sum_{j=1}^N t_j \middle| t_j \in T^i \right\}, \\ \bigcup_{i=1}^n dT^i &= T. \end{aligned} \tag{77}$$

Taking $n \to \infty$, it converges to

$$\begin{aligned} P(x_1, x_2, \cdots, x_N|T) &= \int_T \lambda_N(dT_N) \\ &= \int_T \mu_1(dT)\mu_2(dT)\cdots\mu_N(dT). \end{aligned} \tag{78}$$

Therefore, based on the commonality as convolution, we derive whole orders of the joint distribution necessary for the calculation of known complexity measures. In a general form, any complexity measure incorporating the information of a joint distribution can be described as the function of convolution $G^{-1}(Q(\lambda))$, following the formalization of section 2.1.

Commonality order is also accessible to existing algorithms that extract the total order of system elements, such as Dulmage-Mendelsohn decomposition [41] and phylogenetic tree analyses [42]. Although the calculation of joint distributions of all orders out of matrix data generally confronts exponential computational time, total order based on partial combinatorics and statistical testing with known distribution of p-value can provide a quick evaluation of matching on the results from different algorithms. The pair-wise and triplet order algorithms of N observations can be processed with $\mathcal{O}(N^2)$ and $\mathcal{O}(N^3)$, respectively, similar to the range of most other ranking algorithms that is based on low-order statistics. The comparison between N total orders of commonality requires only second degree polynomial time $\mathcal{O}(N^2)$ (38). Taking such partial optimization and algorithm-wise comparison of performance into account, as an extensive Bayesian estimator including human of section 4.2, deep learning model with the use of massive parallel machine learning can be structurally effective for an interactive recombination of an estimation model based on human feedback

In order to effectively attain criticality in citizen science where knowledge acquisition, transfer, and control are optimized through self-organization, we need to reach a collective intelligence that is distributed in a parallel way both in our subjective mind and objective reality. The cost of data-driven science sometimes depends on the overly weighted objective measurement for complete modelling, which can also hinder the agility of taking actions, and opportunity of effective interaction through internal observation [3]. As explored in this article, if there exist natural laws extended in our collective intelligence—much like the physical law in objective nature—we may count on such topological structure, and it may be possible to take effective guidance through partial and distributed observation. Such a way to organize collective intelligence among independent and parallel activity producers could be considered as a social-environmental expansion of the "intelligence without representation", which is based on the direct interface to the world through perception and action, rather than comprehensive representation of knowledge isolated from the environment [43]. Data acquisition needs to generate potentially effective action strategies, or the **affordance** under global management principles, instead of modelling the phenomena without essential intervention of actors [44]. This can be described as the **data-affordance science** in contrast to exhaustive data-driven science, in which we substantially depend on the emergent topological structure of inter-subjective objectivity to take decisions in real time, represented at the intersection of the human mind, computation and natural phenomena. The **buoy-anchor-raft** model developed as a mutual framework can provide a theoretical basis that expands external observation of conventional science to internal observation necessary for the management and knowledge extraction as a **data-affordance** science [5] [27]. As a cumulative effect of synergistic efficiency, observation and data processing could diminish within a computable time scale by implicitly augmenting the knowledge representation incorporated into actual action principles. With measurement-action unity as a process of **affordance** in both data and reality, cost-effective interface and human-dependable system could be realized within the framework of internal observation, as a crucial premise for sustainable solution. The edge of criticality for a successful citizen science, in terms of its nature and resource restriction, could find its limits neither in our internal mind nor external world, but on the topology of these interactions.

The author declares no conflict of interest.

Appendix

Proof of Theorem 1. Let us formulate Eq.(12) as $I = F^{-1}(Q(x))$. As I is an epimorphism but not necessarily a monomorphism, its inverse function generally retrieves a larger subset of conditions including $Q(x)$:

$$F(I) \supseteq Q(x). \tag{A1}$$

Recursively defining $Q'(x)$ by specifying the value of I as

$$Q'(x) = S_R^{-1}\left[\{x \in \mathbb{X} | I = const.\}\right], \tag{A2}$$

one obtains the inverse function that brings us back exactly to the comprehensive search condition, $F'(I) = Q'(x)$.

Now, we consider the epimorphism $H : \{Q(x)\} \to \{Q'(x)\}$ with its right-sided inverse as $H^{-1} : \{Q'(x)\} \to \{Q(x)\}$ and $H^{-1} \circ H \circ Q(x) = Q(x)$. We set $Q'(x) = H \circ Q(x)$, which gives $F'(I) = H \circ Q(x)$, then $H^{-1} \circ F'(I) = Q(x)$. Next, we consider I' such that $F'(I') = Q(x)$. By resolving $F' \circ F''(I) = H^{-1} \circ F'(I)$ with respect to $F'' : \mathbb{R} \mapsto \mathbb{R}$, we obtain

$$I' = F''(I), \tag{A3}$$

then

$$Q(x) = F' \circ F''(I) = F'(I'), \tag{A4}$$

which shows coincidence between F' and G with exclusively selective complexity measure I'. The exact construction of Q', F', H, and F'' depends on the exhaustive computation process, whose computational complexity is characterized in the section 3. □

Proof of Theorem 2. From Tonelli's theorem,

$$\begin{aligned}
\lambda_N(r_N) &:= \mu_1 * \mu_2 * \cdots * \mu_i * \cdots * \mu_N(r) \\
&= \int_{\mathbb{R}} \cdots \left(\int_{\mathbb{R}} \cdots \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1} \left(\Lambda \sum_{i=1}^{N} x_i \middle| x_i \in r \right) \mu_1(dx_1) \right) \mu_2(dx_2) \right) \cdots \mu_i(dx_i) \right) \cdots \mu_N(dx_N) \\
&= \left(\int_{\mathbb{R}} \mathbf{1}\left(x_1 | x_1 \in r\right) \mu_1(dx_1) \right) \left(\int_{\mathbb{R}} \mathbf{1}\left(x_2 | x_2 \in r\right) \mu_1(dx_2) \right) \cdots \\
&\qquad \cdots \left(\int_{\mathbb{R}} \mathbf{1}\left(x_i | x_i \in r\right) \mu_1(dx_i) \right) \cdots \left(\int_{\mathbb{R}} \mathbf{1}\left(x_N | x_N \in r\right) \mu_1(dx_N) \right) \\
&= \prod_i^N q_i(r).
\end{aligned} \tag{A5}$$

□

Proof of Theorem 3. The central limit theorem with Lindeberg's condition assures the following convergence as the sampling number $N' \to \infty$ and the number of distribution $N \to \infty$:

$$\sum_{j=1}^{N'} \sum_{i=1}^{N} \frac{x_{ij}}{N'} \to \int_{\mathbb{R}} \mathcal{N}(\nu'_N, \sigma'^2_N) m(dx), \tag{A6}$$

where the variables $x_{ij} \in X_i = \{x_{i1}, \cdots, x_{iN'}\}$ follow independent distributions $p(X_i)$, $X_i \in \{X_1, \cdots, X_N\}$, with finite mean $\alpha'_i = \sum_{j=1}^{N'} \frac{x_{ij}}{N'}$ and variance $\beta'^2_i = \sum_{j=1}^{N'} \frac{x_{ij}^2}{N'} - \alpha'^2_i$ taken over N' samples, and

$$\nu'_N = \sum_{i=1}^{N} \alpha'_i, \tag{A7}$$

$$\sigma'^2_N = \sum_{i=1}^{N} \beta'^2_i. \tag{A8}$$

Based on the central limit theorem, we consider the numerical convergence of $\lambda_N(r_N)$ in a way accessible to $r_S \subseteq r_N$. The convolution $\lambda_N(r_N)$ represents infinite random sampling of $x \in r_N := \left\{ \Lambda \sum_{k=1}^{N} x_k \middle| x_k \in r \right\}$ at the limit of $N' \to \infty$, from N independent distributions $\{\mu_i(x)|x \in r, i = 1, \cdots, N\}$ as the population distributions with finite mean α_i and variance β_i^2 as follows:

$$\begin{aligned} \alpha_i &= \int_{\mathbb{R}} \mathbf{1}\,(x|x \in r)\, x\mu_i(dx), \\ \beta_i^2 &= \int_{\mathbb{R}} \mathbf{1}\,(x|x \in r)\, x^2\mu_i(dx) - \alpha_i^2, \end{aligned} \tag{A9}$$

where each mean and variance is bounded within the total variation of r as

$$\inf(r) \le \alpha_i \le \sup(r), \tag{A10}$$

$$\beta_i \le \frac{\sup(r) - \inf(r)}{2} = \frac{||r||}{2}. \tag{A11}$$

If $\mu_i(x \in r)$ are finite measures, we obtain the following from the central limit theorem of independent distributions with finitely bounded mean and variance:

$$\begin{aligned} \lambda_N(r_N) &\to \int_{\mathbb{R}} \mathbf{1}\,(x|x \in r_N)\, \mathcal{N}(\Lambda\nu_N, \Lambda^2\sigma_N^2) m(dx) \\ &\quad \times \int_{\mathbb{R}^N} \mathbf{1}\left(\Lambda \sum_{i=1}^{N} x_i \middle| x_i \in r \right) \mu_1(dx_1) \cdots \mu_i(dx_i) \cdots \mu_N(dx_N) \\ &= \int_{\mathbb{R}} \mathbf{1}\,(x|x \in r_N)\, \mathcal{N}(\Lambda\nu_N, \Lambda^2\sigma_N^2) m(dx) \times \prod_{i}^{N} q_i(r), \end{aligned} \tag{A12}$$

where

$$\begin{aligned} \nu_N &= \sum_{i=1}^{N} \alpha_i, \\ \sigma_N^2 &= \sum_{i=1}^{N} \beta_i^2, \end{aligned} \tag{A13}$$

Peer-reviewed version available at *Entropy* **2017**, *19*, , 181; doi:10.3390/e19040181

which coincides with (33) as $N \to \infty$. In (A12), the term $\prod_{i=1}^{N} q_i(r)$ serves as the overall normalisation factor since $q_i(r)$ is not necessarily normalized as a probability distribution with total probability 1. Since the convolution is replaced by the integral of normal distribution with single variable, by restricting on arbitrary subset $r_s \subseteq r_N$, we obtain the theorem (32):

$$\begin{aligned} \lambda_N(r_s) &\to \int_{\mathbb{R}} \mathbf{1}\,(x|x \in r_s)\,\mathcal{N}(\Lambda \nu_N, \Lambda^2 \sigma_N^2) m(dx) \times \prod_i^N q_i(r) \\ &= \int_{r_s} \mathcal{N}(\Lambda \nu_N, \Lambda^2 \sigma_N^2) m(dx) \times \prod_i^N q_i(r). \end{aligned} \tag{A14}$$

In case $\mu_i(x \in r)$ include infinite measures that do not guarantee the above convergence, ${}^{\exists}x \in r$, such that $\mu_i(x) = \infty$, though

$$m(\{x \in r | \mu_i(x) = \infty\}) = 0. \tag{A15}$$

Because, in the opposite case, $m(\{x \in r|\mu_i(x) = \infty\}) > 0$, $\mu_i(r) = q_i(r) = \infty$, which contradicts the definition (24). Since infinite measures could only appear within a countable set of zero Lebesgue measure,

$$\mu_i(\{x \in r | \neg \text{Theorem 3 }\}) = 0, \tag{A16}$$

which means for almost every $x \in r_s$, the theorem holds. □

Proof of Theorem 4. The null hypothesis can be represented as a random order distribution, in which $M = {}_N C_2$ pairs of N observations are susceptible to generate an I-N compromise between I and II. Choose an arbitrary commonality order I and consider the null hypothesis distribution of II.

With respect to an arbitrary pair (i, j) out of N observations, all permutations in $\mathcal{G}_N$ can be divided into 2 sets $\mathcal{H}_{\text{I-I}}$ and $\mathcal{H}_{\text{I-N}}$, which correspond to those generating I-I matching and I-N compromise, respectively:

$$\mathcal{H}_{\text{I-I}} := \{g_{\text{I-I}} \in \mathcal{G}_N | g_{\text{I-I}}(i) = i', g_{\text{I-I}}(j) = j', 1 \le i' < j' \le N\}, \tag{A17}$$

$$\mathcal{H}_{\text{I-N}} := \{g_{\text{I-N}} \in \mathcal{G}_N | g_{\text{I-N}}(i) = j', g_{\text{I-I}}(j) = i', 1 \le i' < j' \le N\}. \tag{A18}$$

Here, $\mathcal{H}_{\text{I-I}}$ and $\mathcal{H}_{\text{I-N}}$ are not groups but the subsets of the same size,

$$\#(\mathcal{H}_{\text{I-I}}) = \#(\mathcal{H}_{\text{I-N}}) = \frac{1}{2}\#(\mathcal{G}_N), \tag{A19}$$

because

$$\mathcal{H}_{\text{I-N}} = g_{i'j'} \circ \mathcal{H}_{\text{I-I}}, \tag{A20}$$

$$\mathcal{H}_{\text{I-I}} \cup \mathcal{H}_{\text{I-N}} = \mathcal{G}_N, \tag{A21}$$

$$\mathcal{H}_{\text{I-I}} \cap \mathcal{H}_{\text{I-N}} = \emptyset, \tag{A22}$$

where for $k = 1, \cdots, N$,

$$g_{i'j'}(k) := \begin{cases} j' & \text{if} \quad k = i', \\ i' & \text{if} \quad k = j', \\ k & \text{else.} \end{cases} \tag{A23}$$

Then with respect to the random permutation, the probability p that each pair from N observations will be judged as I-I matching is given by:

$$p = \frac{\#(\mathcal{H}_{\text{I-I}})}{\#(\mathcal{G}_N)} = 0.5, \tag{A24}$$

which leads to the general probability of the occurence number of I-I matching $(k_{\text{I-I}} \geq 1)$ follow binomial distribution with parameters $M = {}_NC_2$ and p.

Note that the binomial distribution can be approximated to a normal distribution with $N \geq 7$ in this case, according to the condition of the mean value $Mp > 5$ and variance $Mp(1-p) > 5$. □

Proof of Theorem 5. Take $n > m \in \mathbf{N}$ and consider the database $\mathbb{X}$, $\#(\mathbb{X}) = n$, in which we divide m observations with $k = \left\lfloor \frac{n}{m} \right\rfloor$ elements and these intersections as commonality structure. $\lfloor \cdot \rfloor$ is a floor function.

As the cardinality of rational number is $\aleph_0$, any positive common fraction, or $\mathbb{N}^2$, can find unique correspondence to $\mathbb{N}$. Now, for an arbitrary $k = \left\lfloor \frac{n}{m} \right\rfloor$, $\exists n'$, such that $k < \left\lfloor \frac{n'}{m} \right\rfloor$. (For example, take $n' = m \left\lceil \frac{n}{m} \right\rceil$ with ceiling function $\lceil \cdot \rceil$.) Since $k \in \mathbb{N}$, for simplicity, let us consider the correspondence $n = km$ for any $k, m \in \mathbb{N}$. With the use of Cantor's pairing function $\langle \cdot, \cdot \rangle : \mathbb{N}^2 \mapsto \mathbb{N}$, we obtain the unique counting natural number $\langle k, m \rangle$ for all pairs of (k, m):

$$\langle k, m \rangle := \frac{1}{2}(k+m)(k+m+1) + m. \tag{A25}$$

As $n = km$ contains permutational symmetry with respect to k and m, the uniqueness does not hold for $\langle k, m \rangle \mapsto n$, though from the inverse function of $\langle \cdot, \cdot \rangle$,

$$\lim_{\langle k,m \rangle \to \infty} k = \infty, \tag{A26}$$

$$\lim_{\langle k,m \rangle \to \infty} m = \infty, \tag{A27}$$

$$\lim_{\langle k,m \rangle \to \infty} km = \lim_{\langle k,m \rangle \to \infty} n = \infty. \tag{A28}$$

As $n \to \infty$ is equivalent with either $k \to \infty$ or $m \to \infty$,

$$\lim_{n \to \infty} \langle k, m \rangle = \infty, \tag{A29}$$

which results in

$$\lim_{n \to \infty} k = \infty, \tag{A30}$$

$$\lim_{n \to \infty} m = \infty. \tag{A31}$$

Taking $n = \#(\mathbb{X})$, $m = O(r)$ and $k = \#(\{r|O(r) = m\})$ gives the theorem. □

Proof of Theorem 6. From the definition of $\Delta(i)$, when there is no diminution of data or equivalently $\lambda_k(\mathbb{X}) > 0$ for all $k \in \left\{1, \cdots, N' = \left\lfloor \frac{N}{2} \right\rfloor\right\}$,

$$\begin{aligned} \mathcal{O}\left(\prod_i^{N'} \Delta(i)\right) &= \frac{\mathcal{O}(N^2)}{\mathcal{O}(1)} \cdot \frac{\mathcal{O}(N^3)}{\mathcal{O}(N^2)} \cdot \frac{\cdots}{\mathcal{O}(N^3)} \cdots \frac{\mathcal{O}(N^{N'-1})}{\cdots} \frac{\mathcal{O}(N^{N'})}{\mathcal{O}(N^{N'-1})} \\ &= \mathcal{O}(N^{N'}) \\ &= \sqrt[d]{\mathcal{O}(N^{dN'})}. \end{aligned} \tag{A32}$$

As the product monotonically decreases with respect to the decrease of each element, the above relation gives the upper bound. Sorting time of N elements is usually given by $\mathcal{O}(N^2)$, $d = 2$, and can be generalized to algorithms with polynomial order $d > 0$. □

Proof of Theorem 7. From

$$\Delta'(i,j) = \Delta(i)\Delta(j), \tag{A33}$$

we directly obtain

$$\begin{aligned} \mathcal{O}\left(\prod_{(i,j)}^{(N',M')} \Delta'(i,j)\right) &= \mathcal{O}\left(\prod_i^{N'} \Delta(i)\right) \mathcal{O}\left(\prod_j^{M'} \Delta(j)\right) \\ &= \mathcal{O}(N^{N'} M^{M'}) \\ &= \sqrt[d]{\mathcal{O}([N^{N'} M^{M'}]^d)}. \end{aligned} \tag{A34}$$

□

Proof of Theorem 8. The condition (45) can be translated into the following with respect to the data sparseness u:

$$\mathcal{O}([{}_N C_{\dim(\{\mathbb{X}_N|O(\mathbb{X})=N'\})} {}_M C_{\dim(\{\mathbb{X}_M|O(\mathbb{X})=M'\})}]^d) \leq \mathcal{O}([N+M]^c), \tag{A35}$$

$$\mathcal{O}([u_N C_{N'} \cdot u_M C_{M'}]^d) \leq \mathcal{O}([N+M]^c), \tag{A36}$$

$$\mathcal{O}([u^2 N^{N'} M^{M'}]^d) \leq \mathcal{O}([N+M]^c). \tag{A37}$$

Expressed as the order of computational time on both sides of formula without $\mathcal{O}(\cdot)$ for simplicity,

$$[u^2 N^{N'} M^{M'}]^d \quad \leq \quad [N+M]^c, \tag{A38}$$

and taking logarithmic scale,

$$d[\log(u^2 N^{N'} M^{M'})] \leq c \log L, \tag{A39}$$

$$\frac{1}{2} \sum_k^{\{N,M\}} \left\lfloor \frac{k}{2} \right\rfloor \log k \leq \frac{c}{2d} \log L - \log u. \tag{A40}$$

We consider the application of Chebyshev's inequality on the left side, such that

$$\frac{1}{2}\sum_{k}^{\{N,M\}}\left\lfloor\frac{k}{2}\right\rfloor\cdot\frac{1}{2}\sum_{k}^{\{N,M\}}\log k\leq\frac{1}{2}\sum_{k}^{\{N,M\}}\left\lfloor\frac{k}{2}\right\rfloor\log k. \tag{A41}$$

Since $\left\lfloor\frac{k}{2}\right\rfloor\Big/\frac{k}{2}\to 1$ as $k\to\infty$ and removing constant coefficient $\frac{1}{2}$, evaluation of asymptotic behaviour of (A41) can be derived essentially for the left side from $f_l(N,M)$ and the right side from $f_r(N,M)$ defined as follows,

$$\begin{array}{rcl} f_l(N,M) & := & (N+M)(\log N+\log M), \\ f_r(N,M) & := & N\log N+M\log M, \end{array} \tag{A42}$$

with which (A41) is described as

$$\frac{1}{2}f_l(N,M)\leq f_r(N,M). \tag{A43}$$

As $N,M\to\infty$ that becomes dominant as $L\to\infty$,

$$\begin{array}{rcl} f_l(N,M) & \leq & \mathcal{O}((N+M)\log(N+M)), \\ f_r(N,M) & \leq & \mathcal{O}((N+M)\log(N+M)), \end{array} \tag{A44}$$

since ${}^{\exists}D>0$, ${}^{\exists}C>0$, such that $N,M\geq D$ then $f_l(N,M),f_r(N,M)\leq C\cdot[(N+M)\log(N+M)]$. This condition holds with $D\geq 1, C\geq 2$ for both $f_l(N,M)$ and $f_r(N,M)$. Note that although explicit inequality between $f_l(N,M)$ and $f_r(N,M)$ exists in (A43), these converge to the same asymptotic order $\mathcal{O}(L\log L)$ for all N and M, because as $L\to\infty$,

$$\begin{array}{rcccl} 1 & \leq & \dfrac{f_r(N,M)}{f_l(N,M)} & \leq & 2, \\ \dfrac{1}{2} & \leq & \dfrac{f_l(N,M)}{L\log L} & \leq & 1, \end{array} \tag{A45}$$

and

$$\frac{f_r(N,M)}{L\log L}\to 1, \tag{A46}$$

which remain within the ranges of multiplication with constant. The relations (A45) and (A46) can be proved by examining the minimum and maximum values of $\frac{f_r(N,M)}{f_l(N,M)}$, $\frac{f_l(N,M)}{L\log L}$ and $\frac{f_r(N,M)}{L\log L}$. By considering with the range of $1\leq N\leq\frac{L}{2}$ from the symmetry between N and M ($M=L-N$), we derive the following monotonicity conditions with respect to N,

$$\frac{\partial f_l(N,M)}{\partial N}\geq 0, \tag{A47}$$

$$\frac{\partial f_r(N,M)}{\partial N}\leq 0, \tag{A48}$$

from which we obtain the minimum value of $\frac{f_r(N,M)}{f_l(N,M)}$ at $N = \frac{L}{2}$,

$$\frac{f_r\left(\frac{L}{2},\frac{L}{2}\right)}{f_l\left(\frac{L}{2},\frac{L}{2}\right)} = 1, \tag{A49}$$

the maximum value of $\frac{f_r(N,M)}{f_l(N,M)}$ at $N = 1$,

$$\frac{f_r(1,L-1)}{f_l(1,L-1)} = 2 \cdot \frac{L-1}{L} \to 2, \tag{A50}$$

the minimum value of $\frac{f_l(N,M)}{L\log L}$ at $N = 1$,

$$\frac{f_l(1,L-1)}{L\log L} = \frac{1}{2}\frac{\log(L-1)}{\log L} \to \frac{1}{2}, \tag{A51}$$

the maximum value of $\frac{f_l(N,M)}{L\log L}$ at $N = \frac{L}{2}$,

$$\frac{f_l\left(\frac{L}{2},\frac{L}{2}\right)}{L\log L} = \frac{\log L - \log 2}{\log L} \to 1, \tag{A52}$$

the minimum value of $\frac{f_r(N,M)}{L\log L}$ at $N = \frac{L}{2}$,

$$\frac{f_r(1,L-1)}{L\log L} = \frac{L-1}{L}\frac{\log(L-1)}{\log L} \to 1, \tag{A53}$$

and the maximum value of $\frac{f_r(N,M)}{L\log L}$ at $N = 1$,

$$\frac{f_r\left(\frac{L}{2},\frac{L}{2}\right)}{L\log L} = \frac{\log L - \log 2}{\log L} \to 1, \tag{A54}$$

$$\tag{A55}$$

with the associated convergence as $L \to \infty$. Numerical observation of the convergence between $f_l(N,M)$, $f_r(N,M)$ and $L\log L$ is given in Fig. 8 (a).

As it converges to the same asymptotic behaviour $\mathcal{O}((N+M)\log(N+M))$ on both sides of (A43), we apply the left side of Chebyshev's inequality $f_l(N,M)$ to (A40), which gives asymptotical relation

$$\begin{aligned} \frac{1}{8}f_l(N,M) &\leq \frac{c}{2d}\log L - \log u, \\ u &\leq L^{\frac{c}{2d}} N^{-\frac{L}{8}} (L-N)^{-\frac{L}{8}}, \end{aligned} \tag{A56}$$

where coefficient $\frac{1}{8}$ is derived from the relation (A41) including the effect of transformation $N' = \left\lfloor \frac{N}{2} \right\rfloor$ and $M' = \left\lfloor \frac{M}{2} \right\rfloor$. As $L \to \infty$ and taking the sum over N, it converges to the theorem:

$$\begin{aligned} Lu &\le \sum_{N=1}^{L-1} L^{\frac{c}{2d}} N^{-\frac{L}{8}} (L-N)^{-\frac{L}{8}}, \\ u &\le f * f(L). \end{aligned} \tag{A57}$$

Numerical observation of the proof is given in Fig. 8 (b). □

Proof of Theorem 9. We consider the Θ coordinates of $P_o(X)$ as Θ_o, which constitutes the e-flat geodesic $\Theta(w) = \{\theta_i(w)\}$ between $P_s(X)$ and $P_o(X)$ as

$$\Theta(w) := w\Theta_s + (1-w)\Theta_o. \tag{A58}$$

The tangent vector T^e of the e-geodesic is expressed as

$$T^e = \sum_{i=1}^{n} \frac{d}{dw}\theta_i(w)\frac{\partial}{\partial\theta_i} = \sum_{i=1}^{n}\{\theta_i(P_s(X)) - \theta_i(P_o(X))\}\frac{\partial}{\partial\theta_i}, \tag{A59}$$

and the tangent vector T^m of the m-geodesic H_o as

$$T^m = \sum_{i=1}^{n} \frac{d}{dw}\eta_i(P_o(X))\frac{\partial}{\partial\eta_i} = \sum_{i=1}^{n}\{\eta_i(P_a(X)) - \eta_i(P_l(X))\}\frac{\partial}{\partial\eta_i}. \tag{A60}$$

Then the inner product $< T^e, T^m >$ of these tangent vectors at $P_o(X)$ is expressed as

$$< T^e, T^m > = \sum_{i=1}^{n}\{\theta_i(P_s(X)) - \theta_i(P_o(X))\}\{\eta_i(P_a(X)) - \eta_i(P_l(X))\}, \tag{A61}$$

since from the duality of the coordinates in (60),

$$\left\langle \frac{\partial}{\partial\theta_i}, \frac{\partial}{\partial\eta_j} \right\rangle = \begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{else.} \end{cases} \tag{A62}$$

As $P_a(X)$, $P_l(X)$ and $P_o(X)$ are aligned on the m-geodesic, the relation (A61) can be translated to

$$\begin{aligned} < T^e, T^m > &= \sum_{i=1}^{n}\{\theta_i(P_s(X)) - \theta_i(P_o(X))\}\{\eta_i(P_a(X)) - \eta_i(P_o(X))\} \cdot C_1, \\ < T^e, T^m > &= \sum_{i=1}^{n}\{\theta_i(P_s(X)) - \theta_i(P_o(X))\}\{\eta_i(P_l(X)) - \eta_i(P_o(X))\} \cdot C_2, \end{aligned} \tag{A63}$$

with some constant C_1 and C_2.

Now, from the definition of $\nabla^{(\alpha)}$-divergence (65) and its dual divergence (66), the Pythagorean relations between Kullback-Leibler divergences are expressed as

$$
\begin{aligned}
& D^m(P_a : P_o) + D^m(P_o : P_s) - D^m(P_a : P_s) \\
= \; & D^e(P_o : P_a) + D^e(P_s : P_o) - D^e(P_s : P_a) \\
= \; & \sum_{i=1}^{n} \{\theta_i(P_s(X)) - \theta_i(P_o(X))\}\{\eta_i(P_a(X)) - \eta_i(P_o(X))\} \cdot (-1) \\
= \; & -\frac{1}{C_1} < T^e, T^m >, \\
& D^m(P_l : P_o) + D^m(P_o : P_s) - D^m(P_l : P_s) \\
= \; & D^e(P_o : P_l) + D^e(P_s : P_o) - D^e(P_s : P_l) \\
= \; & \sum_{i=1}^{n} \{\theta_i(P_s(X)) - \theta_i(P_o(X))\}\{\eta_i(P_l(X)) - \eta_i(P_o(X))\} \cdot (-1) \\
= \; & -\frac{1}{C_2} < T^e, T^m > .
\end{aligned} \tag{A64}
$$

When orthogonality holds between the e- and m- geodesic, $< T^e, T^m >= 0$ for (A63), which proves the Pythagorean relations from (A64).

Finally, we prove that $P_o(X)$ satisfies the minimum condition (68). By considering $P_{o'}(X)$ with a parameter $w' \neq w$ as

$$
H_{o'} := w' H_a + (1 - w') H_l, \tag{A65}
$$

we obtain the Pythagorean relation

$$
\begin{aligned}
D^m(P_{o'} : P_s) &= D^m(P_{o'} : P_o) + D^m(P_o : P_s), \\
D^e(P_s : P_{o'}) &= D^e(P_o : P_{o'}) + D^e(P_s : P_o).
\end{aligned} \tag{A66}
$$

Since $D^m(P_{o'} : P_o) = D^e(P_o : P_{o'}) \geq 0$ from the definition of divergence, $D^m(P_{o'} : P_s) \geq D^m(P_o : P_s)$ and $D^e(P_s : P_{o'}) \geq D^e(P_s : P_o)$ hold, which means $P_o(X)$ is a stationary point giving the minimum value with respect to $D^m(\cdot : P_s) = D^e(P_s : \cdot)$, on the m-geodesic between $P_a(X)$ and $P_l(X)$. Note that the theorem also holds when $\sum_X P(X)$ takes arbitrary finite values other than 1. □

References

1. Schwab, K. *The Fourth Industrial Revolution*. Crown Business: New York, USA, 2017
2. https://www.usanpn.org/natures_notebook
3. Funabashi, M.; Hanappe, P.; Isozaki,T.; Maes, A.M.; Sasaki, T.; Steels, L.; Yoshida, K. Foundation of CS-DC e-Laboratory: Open Systems Exploration for Ecosystems Leveraging. *First Complex Systems Digital Campus World E-Conference 2015, Springer Proceedings in Complexity*, Springer International Publishing Switzerland: Cham, Switzerland, 2017; 351-374.
4. Funabashi, M. Open Systems Exploration: An Example with Ecosystems Management. *First Complex Systems Digital Campus World E-Conference 2015, Springer Proceedings in Complexity*, Springer International Publishing Switzerland: Cham, Switzerland, 2017; 223-243.
5. Tokoro, M. Open Systems Science: A Challenge to Open Systems Problems. *First Complex Systems Digital Campus World E-Conference 2015, Springer Proceedings in Complexity*, Springer International Publishing Switzerland: Cham, Switzerland, 2017; 213-221.
6. Bak, P. *How Nature Works: The Science of Self-Organized Criticality*. Copernicus: New York, USA, 1996
7. Jensen, H. J. *Self-Organized Criticality*. Cambridge University Press: Cambridge, UK, 1998
8. Takayasu, H.; Sato, A.; Takayasu, A. Stable Infinite Variance Fluctuations in Randomly Amplified Langevin Systems. *Phys. Rev. Lett.* **1997**, *79*, 966.

9. Scanlon, T.M.; Caylor, K.K.; Levin, S.A.; Rodriguez-Iturbe, I. Positive feedbacks promote power-law clustering of Kalahari vegetation. *Nature* **2007**, *449*, 209-212.
10. Gabaix, X. Power Laws in Economics: An Introduction. *J. Econ. Perspect.* **2016**, *30*, 185-206.
11. Alves, L.G.A.; Ribeiroa, H.V.; Lenzi, E.K.; Mendes, R.S. Empirical analysis on the connection between power-law distributions and allometries for urban indicators. *Physica A.* **2014**, *409*, 175-182.
12. Michelucci, P.; Dickinson, J.L. The power of crowds. *Science* **2016**, *351*, 32-33.
13. Hanappe, P.; Dunlop, R.; Maes, A.; Steels, L.; Duval, N. Agroecology: A Fertile Field for Human Computation. *Human Computation* **2016**, *1*, 1-9.
14. Scott, S.L. A modern Bayesian look at the multi-armed bandit. *Appl. Stochastic Models Bus. Ind.* **2010**, *26*, 639-658.
15. Prokopenko, M. *Guided Self-Organization: Inception*. Springer-Verlag: Berlin Heidelberg, Germany, 2014
16. Rekimoto, J.; Nagao, K. The World through the Computer: Computer Augmented Interaction with Real World Environments. *Proceedings of UIST'95* **1995**, 29-36.
17. Funabashi, M. IT-Mediated Development of Sustainable Agriculture Systems: Toward a Data-Driven Citizen Science. *Journal of Information Technology and Application in Education* **2013**, *2(4)*, 179-182.
18. https://www.cbd.int/sp/targets/
19. Funabashi, M. Synecological farming: Theoretical foundation on biodiversity responses of plant communities. *Plant Biotechnology* **2016**, *33*, 213-234.
20. Goodchild, M.F. Citizens as sensors: the world of volunteered geography. *GeoJoumal* **2007**, *69*, 211-221.
21. ISC-PIF (Institut des Systèmes Complexes, Paris Île-de-France). French Roadmap for Complex Systems. ISC-PIF, http://cnsc.unistra.fr/uploads/media/FeuilleDeRouteNationaleSC09.pdf (2009).
22. Solomon, R. C. Subjectivity. In *Oxford Companion to Philosophy*; Honderich, T. Oxford University Press: Oxford, UK, 2005; pp. 900.
23. Gillespie A.; Cornish F. Intersubjectivity: Towards a Dialogical Analysis. *J Theory Soc Behav* **2009**, *40*, 19-46.
24. https://www.galaxyzoo.org/
25. http://www.inaturalist.org/
26. Rowell, D.L. *Soil Science: Methods & Applications*. Wiley: New York, USA, 1994
27. Kitano, H. Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery. *AI MAG* **2016**, *37*.
28. Ioannidis, J. P. Why most published research findings are false. *PLoS Med* **2005**, *2*, e124.
29. http://linkeddata.org
30. Akao, Y. *QFD: Quality Function Deployment - Integrating Customer Requirements into Product Design*. Productivity Press: New York, USA, 2004
31. Hawker, G.A.; Mian, S.; Kendzerska, T.; French, M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res (Hoboken)* **2011**, *63*, 240-252
32. Funabashi, M. Network Decomposition and Complexity Measures: An Information Geometrical Approach. *Entropy* **2014**, *16*, 4132-4167.
33. Walter, R. *Fourier analysis on groups, Interscience Tracts in Pure and Applied Mathematics, No. 12*. Wiley: New York - London, USA - UK, 1962
34. https://commons.wikimedia.org/wiki/File:Symmetrical_5-set_Venn_diagram.svg
35. Funanashi, M. Synthetic Modeling of Autonomous Learning with a Chaotic Neural Network International Journal of Bifurcation and Chaos. *Int. J. Bifurcation Chaos* **2015**, *25*
36. Doya, K.; Ishii, S.; Pouget, A.; Rao, R.P.N. *Bayesian Brain: Probabilistic Approaches to Neural Coding*. The MIT Press: Cambridge, USA, 2007
37. Yoshimura, J.; Clark, C.W. Individual adaptations in stochastic environments. *Evol. Ecol.* **1991**, *5*, 173-192.
38. Harte, J. *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*. Oxford University Press: Oxford, UK, 2011
39. Amari, S.; Nagaoka, H. *Method of information geometry*. American Mathematical Society: Rhode Island, USA, 2007

40. Rao, C.R. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society* **1945**, *37*, 81-91.
41. Murota, K. *Matrices and Matroids for Systems Analysis*. Springer-Verlag: Berlin, Germany, 2000
42. Roy, S.S.; Dasgupta, R.; Bagchi, A. A Review on Phylogenetic Analysis: A Journey through Modern Era. *Computational Molecular Bioscience* **2014**, *4*, 39-45.
43. Brooks, R.A. Intelligence without representation. *Artif. Intell.* **1991**, *47*, 139-159.
44. Gibson, J.J. *The Ecological Approach to Visual Perception*. Houghton Mifflin: Boston, USA, 1979