A Metric to Quantify Subjective Information in Biological Gradient Sensing

Tyler S. Barker
School of Computing
University of Nebraska-Lincoln
Lincoln, USA
tyler.barker@huskers.unl.edu

Peter J. Thomas
Dept. of Math, Appl. Math, & Statistics
Case Western Reserve University
Cleveland, USA
pjthomas@case.edu

Massimiliano Pierobon
School of Computing
University of Nebraska-Lincoln
Lincoln, USA
maxp@unl.edu

***Abstract*—Information theory has been successfully applied to biology with interesting results and applications, ranging from scientific discovery, to system modeling, and engineering. Novel concepts such as semantic and useful information have been proposed to address the peculiarity of biological systems in contrast to Shannon's classical theory. In this paper, the concept of subjective information, previously observed as an emergent property in a simulated biological system with determinate characteristics, is further explored through the proposal of a novel metric for its quantification. This measure is based on a biological system's ability to dynamically sense and react to environmental signals to achieve a goal. The novel metric is validated through the simulation of a computational model that enables its correlation with different strategies for information acquisition from the environment and processing. The obtained results indicate that the proposed measure of subjective information is reliable in quantifying the effectiveness of a biological system's strategy in using information from the environment for its growth and survival.**

***Index Terms*—Information theory, computational simulation of biological cells, chemical reception, mutual information, semantic information, chemotaxis**

I. Introduction

In recent years, the application of information theory to the study of biological systems has emerged as a powerful and versatile approach for understanding the underlying mechanisms governing the intricate processes of life [1]. The ultimate goal is to develop a quantitative understanding of the storage, transmission, sensing, reception and processing of information within biological systems at different biological scales and contexts [2], [3]. Some areas set to directly benefit from this framework are the identification of the functional relationships among biological components [4], [5], the development of models for their behavior [6], and the definition of design and optimization rules to engineer these systems for a plethora of different applications [7]–[15].

Information theory, originally conceived in the context of engineered communication systems, is often applied to biology in conjunction with the concept of "fitness", or the efficiency of a biological system to achieve determinate goals, *e.g.*, individuals' or species' growth and survival [16]. Seminal works have proposed the equivalence in biological systems between the amount of acquired and processed information (mutual information rates) from its environment and growth (expected log population growth rates) [17], advancing the concept that more information available to a living organism translates into a greater advantage in survival and reproduction. According to the *infomax* principle, instead, mostly applied to neuroscience, a biological system aims (and will adapt) to maximize the amount of information from its environment only when it is known to be important for its internal processes [18].

Stemming from these concepts, Shannon's classical information acquisition and processing (mutual information), *i.e.*, the reduction in uncertainty about a transmitted signal, upon receiving a (noisy) version of the signal, has been contrasted to another concept in biological systems, *i.e.*, the reduction of uncertainty *about those components of the system that matter to the receiver.* This distinction leads to the designation of the former as *syntactic* information, and the latter as *semantic information* [19], [20], which is a portion of the total syntactic information that is useful for optimizing the aforementioned fitness of a biological system (given particular system and environment characteristics) [19]. The further concept of *useful* information extends the semantic information to account for a biological system that does not necessarily have an optimal fitness, but acquires and processes information from the environment according to a certain "strategy" (resulting in a certain - possibly non-optimal - fitness) [21].

In our previous work [22], we showed the emergence of another type of information, which we called "subjective", in a simulated biological system with a specific information acquisition and processing strategy and determinate environmental conditions. In this paper, we propose a novel methodology to quantify the subjective information based on the useful information and accounting for a continuum of different strategies. For this, we devised a novel biological system model where different information acquisition and processing strategies are controlled by two different parameters, and we modified our computational model accordingly. We propose a formula to quantify the subjective information and we present numerical results that contrast it with the syntactic information that can be measured in the simulated system. In the scope of this paper, we solely consider the subjective information

This material is based upon work supported by the National Science Foundation under Grants No. CCF-1816969 and DMS-2052109, and by the Oberlin College Department of Mathematics.

contribution from the information acquisition strategy based on chemoreception. When the environment has specific characteristics, the correlation between the subjective information and the growth and survival of the simulated biological system is apparent, while the same cannot be observed for the syntactic information.

The paper is organized as follows. In Sec. II we introduce concepts and assumptions underlying the proposed subjective information model and quantification metric. In Sec. III, we present the system model. Formulas for measuring syntactic and subjective information measures are specified in Sec. IV. The computational model is described in Sec. V. Finally, numerical results based on the computational model simulations are shown in Sec. VI, and conclusions are drawn in Sec. VII.

II. Abstract Model

We abstract a biological system and the environment where it survives and grows according to the diagram shown in Fig. 1. Given a biological system that acquires and processes information from the environment according to a specific strategy to achieve a growth and survival objective, **Useful Information** I_{U} can be defined as the amount of information that contributes to the execution of that strategy. The information from the environment is acquired through an Information Acquisition Strategy from Environmental Signals, *e.g.*, through gradient sensing, and this information is then processed according to an Information Processing Strategy, which ultimately uses the processed information to change the system's Environmental State, *e.g.*, its location in the environment. The latter will impact the Internal State of the biological system, *e.g.*, food resources at the new location that contribute to the growth and survival of the system, and the environmental signals that can be acquired by the system. In general, the biological system can utilize information about its internal state to modulate its information acquisition and information processing strategies.[1] We broadly define **Subjective Information** I_{subj} as the difference in the useful information obtained by (i) a biological system with internal-state-dependent strategies versus (ii) the same system with internal-state-independent strategies.

A. Model Assumptions

Our definition of subjective information (below) relies on the following assumptions [22]:

1) The environment (signals, food distributions) are dynamic (non-constant) over some temporal or spatial domain.
2) The environment provides two or more distinct signals relevant to the system's growth and survival objective.
3) The system's response to environmental information can take into account the system's internal state.
4) The system's internal state is influenced by its environmental state.[2]
5) The system's sensing resources are finite.

[1]We note that the class of systems rigorously considered in [17] does *not* allow for information processing strategies to take into account the dynamically changing needs – *i.e.* the internal state – of the receiver.

[2]It is not directly influenced by information acquisition or processing.

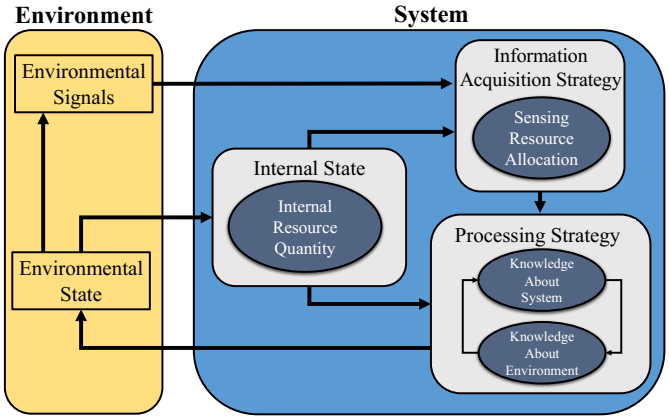

Fig. 1. Block diagram of abstract model for quantifying subjective information.

6) The information acquisition strategy can dynamically manage the system's sensing resources based on the system's internal state.

As shown in Fig. 1, we assume that the biological system has Sensing Resources, *e.g.*, chemoreceptors, that can be allocated according to the information acquisition strategy, and that within the processing strategy the biological system uses the acquired information and the internal state to build a Knowledge about the System and a Knowledge about the Environment. Finally, we assume that an internal state is characterized by an Internal Resource Quantity.

B. Subjective Information Formulation

Suppose we have a biological system composed of individual unicellular organisms that must absorb both of two types of molecules, 'A' and 'B', to survive and grow. We also assume that 'A' and 'B' molecules have different distributions over some environmental domain, where an organism acquires local information on these distributions via chemoreception (gradient sensing) and absorbs them according to their local concentrations. The organism uses the acquired information and its processing strategy to take action, thereby changing its environmental state (location) to maintain a sufficient internal quantity of 'A' and 'B' for growth and survival.

Consider two versions of the system: one where organisms adapt their sensing-resource allocations with respect to their internal stores of 'A' and 'B' ("adapt"), and another where organisms use a fixed strategy, regardless of internal state ("nonadapt"). Relative to a given nonadaptive strategy, the subjective information of the adaptive strategy is:

$$I_{\mathrm{subj}} = I_{\mathrm{U}}(A, B; \mathrm{Action})_{\mathrm{adapt}} - I_{\mathrm{U}}(A, B; \mathrm{Action})_{\mathrm{nonadapt}}, \quad (1)$$

where $I_{\mathrm{U}}(A, B; \mathrm{Action})$ is the "useful information" between the ensemble of the environmental signals A, B and the ensemble of Actions taken by the organism.

III. System Model

To utilize (1) in a concrete example, we specify a biological system satisfying the assumptions in Sec. II-A. Consider a

population of unicellular organisms with length ℓ, whose environmental state is characterized by their center-of-mass positions $\bar{x}$ in a one-dimensional environment (with periodic boundary conditions). Each organism can absorb, and thus acquire information on the distributions of, 'A' and 'B' molecules. Organisms can grow and divide if they acquire enough of both resources, or else perish if they exhaust either resource. Each organism acquires information about the concentrations of 'A' and 'B' molecules at its location through chemoreception. Importantly, each cell has a finite number of receptors, which they can allocate toward sensing either 'A' or 'B'. The receptors detect concentration through a ligand-receptor binding process. Each receptor is bound, independently of the others, with probability $P_c = c/(c+K)$, where c is the 'A' or 'B' concentration at the right or left extreme of the cell $\left(\bar{x} \pm \frac{\ell}{2}\right)$, and K is a dissociation constant. The change in internal resource quantity for either 'A' or 'B' is given by an absorption rate, $\Delta c_{\text{int}} = k\left\{[c]\left(\bar{x}-\frac{\ell}{2}\right)+[c]\left(\bar{x}+\frac{\ell}{2}\right)\right\}/2$, where c_{int} is the internal resource quantity of molecule type c, and k is the absorption coefficient. In addition, the cell consumes 'A' and 'B' at a steady rate given by the metabolic survival cost S, according to the rate law $dc_{\text{int}}/dt = \Delta c_{\text{int}} - S$. If either A_{int} or B_{int} falls to zero, the organism dies. Finally, if both A_{int} and B_{int} exceed a division threshold D, the organism will divide, and split its internal molecule stores equally between its daughter cells. For further details on the model, see [22].

Fig. 2 shows how the model organism acquires and processes information to take action, *i.e.*, to move. Given a set of bound receptors A_b^+ (right) and A_b^- (left) and the number of allocated receptors A_{rec}, we find a normalized signal $A_{\text{sig}} \equiv \frac{A_b^+ - A_b^-}{A_{\text{rec}}/2}$, likewise for 'B'. The A_{sig}, B_{sig} inputs are then interpreted by the processing strategy by multiplying each input by a ratio, respectively $\text{Ratio}_{\text{Proc}}$ or $(1-\text{Ratio}_{\text{Proc}})$, to set the strategy's emphasis on each molecule type. The sum of the resulting quantities is then multiplied by the maximum velocity V_{max} to determine the movement speed and direction. The new allocations of the 'A' and 'B' receptors are also multiplied by a ratio $\text{Ratio}_{\text{Acq}}$ that defines the information acquisition strategy's molecule emphasis. These ratios are determined by a sigmoidal function and a gain that governs the steepness of the switch between an 'A' and 'B' emphasis:

$$\text{Ratio}_{\text{Acq/Proc}} = \frac{1}{1+\exp\left((-\text{Ratio}_{\text{int}}-0.5)\,\text{Gain}_{\text{Acq/Proc}}\right)}, \tag{2}$$

where $\text{Ratio}_{\text{int}}$ is equal to $\frac{A_{\text{int}}}{A_{\text{int}}+B_{\text{int}}}$.

A. Adaptive and Equivalent Fixed-Receptor Strategies

Our class of model systems was chosen to isolate the effects of the information acquisition strategy from the downstream processing strategy. A change in $\text{Gain}_{\text{Proc}}$ creates different processing strategies that may be effective for the organism. For every individual processing strategy there exist many information acquisition strategies controlled by Gain_{Acq}, but these differences in sensing systems only adjust the sensitivity

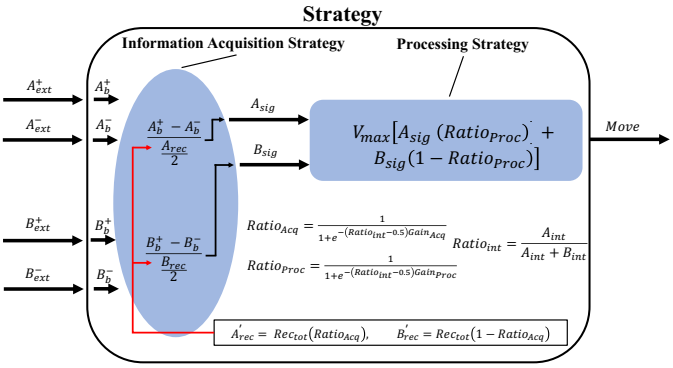

Fig. 2. Model schematic of cell signal acquisition and processing

to a given signal and not the magnitude of the movement performed after processing. Therefore, different information acquisition strategies can result in different amounts of subjective information.

An organisms's information acquisition strategy is defined as *adaptive* if it can dynamically allocate its receptors based on its internal state. Given an adaptive strategy, we define the *equivalent nonadaptive* strategy as having its receptors equally divided among receptor types 'A' and 'B' (*i.e.* with Gain_{Acq} equal to 0), while incorporating the same processing strategy (*i.e.* with the same $\text{Gain}_{\text{Proc}}$) as the corresponding adaptive strategy.

IV. Information Measures

Considering the system model described in Sec. III, the mutual information between the environmental concentrations of molecules 'A' and 'B' and the bound receptors is defined as the syntactic information available to the cell,

$$I_{\text{synt}} = I(A_{\text{ext}}, B_{\text{ext}}; A_{\text{b}}, B_{\text{b}})\,, \tag{3}$$

where $A_{\text{ext}} \in \mathbb{R}_+^2$ and $B_{\text{ext}} \in \mathbb{R}_+^2$ represent the ensemble of external concentration signals seen at the two extremities of the organism. The variables $A_{\text{b}} \in \mathbb{N}^2$ and $B_{\text{b}} \in \mathbb{N}^2$ likewise represent the corresponding bound receptor counts.

We define I_{U} in our model as the mutual information between the environmental concentrations and the movement of the organism given the internal state.[3] This quantity represents the information from the environmental signals that the cell is using to make decisions about where to move,

$$I_{\text{U}}\,|_{\text{strategy}} = \text{E}\left[I\left(A_{\text{ext}}, B_{\text{ext}}; \text{Move} \mid \text{state}\right)\right]_{\text{strategy}}\,, \tag{4}$$

where the random variable Move represents the ensemble of movements produced by the organism in response to the external signals A_{ext} and B_{ext}. The useful information is calculated as an iterated expectation: first we take a conditional expectation, conditioned on the internal state of the organism, i.e. state $\equiv (A_{\text{int}}, B_{\text{in}})$. Subsequently, we average over the internal states resulting from the strategy, $\text{E}[\cdot]_{\text{strategy}}$.

[3]Compare the distinction between information processing in gradient sensing vs. cell movement used in studies of chemotaxis, cf. [23]–[26].

The difference of the useful information I_{U} in a given adaptive strategy, 'adapt' and the corresponding nonadaptive equivalent strategy, 'equiv,' as defined in Sec. III-A, is here expressed as the subjective information I_{subj}:

$$I_{\mathrm{subj}} = I_{\mathrm{U}} \mid_{\mathrm{adapt}} - I_{\mathrm{U}} \mid_{\mathrm{equiv}} . \tag{5}$$

This quantity represents the portion of useful information that the biological system gains by applying an internal-state-dependent information acquisition strategy, independently from the specific information processing strategy. This formula quantifies the informational benefit (towards survival and growth) in the organism's sensing adaptation by allocating receptors sensitive to molecule 'A' or molecule 'B' according to its internal need.

V. Computational Model

The computational model is based on the system model described in Sec. III and enables the calculation of the numerical results in Sec. VI. In this computational model, the organisms' growth rate is determined based on the expected value of the logarithmic difference in population over a given period of time T, namely $\mathrm{E}_{t\in T}\left[\log_2\left(\frac{P_{(t+\Delta t)}}{P_t}\right)\right]$, where P_t is the organisms' population in the biological system at time t. To reduce computational complexity, the organisms' population count during the simulation is constrained within a maximum set threshold. Upon reaching this threshold, the total population count is halved while maintaining the same distribution across the environment. For the final growth rate calculation, the population count is then adjusted to the correct value. Time and space dimensions in the model are sampled according to Δt and Δx, respectively.

In the simulation, we distribute the concentrations of molecules 'A' and 'B' in space according to periodic von Mises distributions expressed as follows (cf. Fig. 3(a)):

$$[C](x) = [C]_{\mathrm{scale}} \frac{\exp\left[\kappa_{\mathrm{C}} \cos\left(\frac{2\pi}{L}\left(x - \mu_{\mathrm{C}}\right)\right)\right]}{2\pi I_0(\kappa_{\mathrm{C}})}, \tag{6}$$

where L is the size of the environment, C is the molecule type (A or B), $\mu_{\mathrm{C}} \in [0, L]$ is the location of the peak concentration, κ_{C} represents the inverse periodic variance of the distribution, $I_0(\kappa_{\mathrm{C}})$ is the Bessel function of order zero, and $[C]_{\mathrm{scale}}$ is a scaling factor.

Each simulation begins with one organism at each of $L/\Delta x$ locations, and is simulated for time 0 to T. The statistics of each organism interaction such as A^+_{ext}, A^+_b, Move, and the overall population is recorded as an ensemble of inputs and outputs. Thus, the input distribution, as seen by the organisms, is different in each simulation. For more information on the computational model and the source code please refer to [27].

A. Information Estimates

Several algorithms are available for estimating mutual information (MI) between multidimensional random variables [28], [29]. For simplicity, we used the following histogram based methodology. Algorithm 1 outlines a procedure for calculating the MI between two variables, X and Y. Here, X is a set of two-dimensional data points, while Y is a set of corresponding data points of dimension one or two. The algorithm uses a procedure, BIN, to bin the data along one or two dimensions with equal bin sizes. BIN takes additional parameters such as edges, which specify the bin boundaries. BIN returns the binned and auxiliary data, a PDF over the bins, and the edges used. The syntactic information estimate is found using Alg 1 as follows,

$$I_{\mathrm{synt}} = \sum_{c\in[A,B]} \mathrm{MI}(X = \left[c^+_{\mathrm{ext}}, c^-_{\mathrm{ext}}\right], Y = \left[c^+_b, c^-_b\right], \mathrm{bins}), \tag{7}$$

with $\mathrm{bins} = 30$ in the simulation.

Algorithm 1 Mutual Information of Multidimensional Data

procedure MI($X = [x_1, x_2], Y = ([y_1]\mathrm{or}[y_1, y_2]), \mathrm{bins}$) ▷ Compute the mutual information between X and Y
  $\mathrm{binned}_X, \mathrm{PDF}_X, \mathrm{edges}_X \leftarrow \textsc{bin}(x_1, x_2, \mathrm{bins})$
  $H_x \leftarrow 0$
  for i in PDF_X(2-D) **do**
    $H_x \leftarrow H_x - \mathrm{PDF}_X[i] \cdot \log_2(\mathrm{PDF}_X[i])$
  end for
  $\mathrm{binned}_Y, \mathrm{PDF}y, \mathrm{edges}_Y \leftarrow \textsc{bin}(y_1, y_2, X, \mathrm{bins})$
  $H_{X|Y} \leftarrow 0$
  for y in PDF_Y **do** ▷ Iterate over i or i,j if PDF_Y is 2-D
    $X \mid y \leftarrow X \mid \mathrm{binned}_Y[y]$
    $\mathrm{binned}_{X|y}, \mathrm{PDF}_{X|y}, \mathrm{edges}_{X|y} \leftarrow \textsc{bin}(X \mid y, \mathrm{edges}_X)$
    $H_{X|y} \leftarrow 0$
    for x in $\mathrm{PDF}_{X|y}$(2-D) **do**
      $H_{X|y} \leftarrow H_{X|y} - \mathrm{PDF}_{xy}[x] \cdot \log_2(\mathrm{PDF}_{xy}[x])$
    end for
    $H_{X|Y} \leftarrow H_{X|Y} + H_{X|y} \cdot PDF_Y[y]$
  end for
  return $H_x - H_{X|Y}$ ▷ Mutual information between X and Y
end procedure

Algorithm 2 outlines a procedure for calculating the useful information I_{U} given a set of data (X, Y) (Y being the corresponding cell velocities), a set of internal resource quantities (internal state) ($A_{\mathrm{int}}, B_{\mathrm{int}}$), the number of bins used in the mutual information calculation, $\mathrm{bins}_{\mathrm{MI}}$, and in binning the internal states, $\mathrm{bins}_{\mathrm{int}}$. $\mathrm{bins}_{\mathrm{MI}}$ is equal to 30 and $\mathrm{bins}_{\mathrm{int}}$ is equal to 10 in the computational model. I_{subj} is then calculated as the difference of I_{U} given the input signals, internal state and the movements of the 'adapt' and 'equiv' types respectively.

VI. Numerical Results

Simulations were performed with the computational model described in Sec. V and the following parameters: length ℓ equal to 1 unit, environmental size $L = 100$, a metabolic cost of S equal to 100, a division threshold of $D = 5 * S$, an absorption constant k equal to 20, a time step Δt equal to 0.01 and a spatial step Δx equal to 0.1. Each simulation was run,

Algorithm 2 Useful Information I_{U}

Require: $X_1 = [x_{1_a}, x_{1_b}]$, $X_2 = [x_{2_a}, x_{2_b}]$ and $Y = ([y_1, y_2])$

procedure $\mathrm{UI}(X_1, X_2, Y, A_{\mathrm{int}}, B_{\mathrm{int}}, \mathrm{bins}_{\mathrm{MI}}, \mathrm{bins}_{\mathrm{int}})$
  $\mathrm{Data} \leftarrow (X_1, X_2, Y)$
  $\mathrm{bin}_{XY}, \mathrm{PDF}_{XY}, \_ \leftarrow \mathrm{BIN}(A_{\mathrm{int}}, B_{\mathrm{int}}, \mathrm{Data}, \mathrm{bins}_{\mathrm{int}})$
  $\mathrm{UI} \leftarrow 0$
  for int in bin_{XY} **do** ▷ int is internal state index
    $X_{1_{|\mathrm{int}}} \leftarrow \mathrm{Data}[X_1] \mid \mathrm{int}$
    $X2_{|\mathrm{int}} \leftarrow \mathrm{Data}[X_2] \mid \mathrm{int}$
    $y_{1_{|\mathrm{int}}}, y_{2_{|\mathrm{int}}} \leftarrow \mathrm{Data}[Y] \mid \mathrm{int}$
    $\mathrm{MI}_{X_{1_{|\mathrm{int}}}} \leftarrow \mathrm{MI}(X_{1_{|\mathrm{int}}}, y_{1_{|\mathrm{int}}}, \mathrm{bins}_{\mathrm{MI}})$
    $\mathrm{MI}_{X_{2_{|\mathrm{int}}}} \leftarrow \mathrm{MI}(X_{2_{|\mathrm{int}}}, y_{2_{|\mathrm{int}}}, \mathrm{bins}_{\mathrm{MI}})$
    $\mathrm{UI} \leftarrow \mathrm{UI} + (\mathrm{MI}_{X_{1_{|\mathrm{int}}}} + \mathrm{MI}_{X_{2_{|\mathrm{int}}}}) \cdot \mathrm{PDF}_{XY} \mid \mathrm{int}$
  end for
  return UI ▷ Useful information between X and Y
end procedure

starting with 1000 cells, for time $T = 20$ with $\mathrm{Gain}_{\mathrm{Acq}}$ and $\mathrm{Gain}_{\mathrm{Proc}}$ in $[-50, 50]$. Each combination of values for $\mathrm{Gain}_{\mathrm{Acq}}$ and $\mathrm{Gain}_{\mathrm{Proc}}$ is color-coded as shown in Fig. 3(h).

In Fig. 3 we show the resulting syntactic and subjective information for simulations performed with two different von-Mises-distributed environments. The simulation of the computational model used 'A' and 'B' molecule concentration profiles with two von Mises distributions 180° (μ_{c} equal to 25 and 75) and 0° apart (μ_{c} both equal to 50), $[C]_{\mathrm{scale}}$ equal to 200, κ_{C} equal to $L/(20\pi)$. The correlation coefficient R is displayed for each plot to see how subjective information correlates with population growth. The identical von Mises distribution is used here to show a situation where subjective information would be identically zero.[4]

The results support the notion of subjective information in the model as the organisms tend to display a higher growth rate with larger subjective information in the disjoint distribution case, resulting in a higher correlation compared to that of syntactic information. As can be seen in Fig. 3(g) this result also holds true for a larger dissociation coefficient K. Fig. 3 (d) and (f) report the cases where there is no subjective information. In this case, there is no correlation with the subjective information values and a larger correlation with the syntactic information values.

The I_{subj} measure for the equivalent organism strategies in Fig. 3 (f) are all close to zero, as expected. Because these cells exhibit no diversity in their input signals between 'A' or 'B' there is no gain in useful information. The set of these organisms does, however, have a variety of growth rates. This variation arises from the different ways the systems process the signals available to them. Having subjective information does not imply that a biological system is optimized for growth

[4]Strictly speaking the $\mu_{\mathrm{A}} = \mu_{\mathrm{B}}$ case violates Assumption 2 in Sec. II-A.

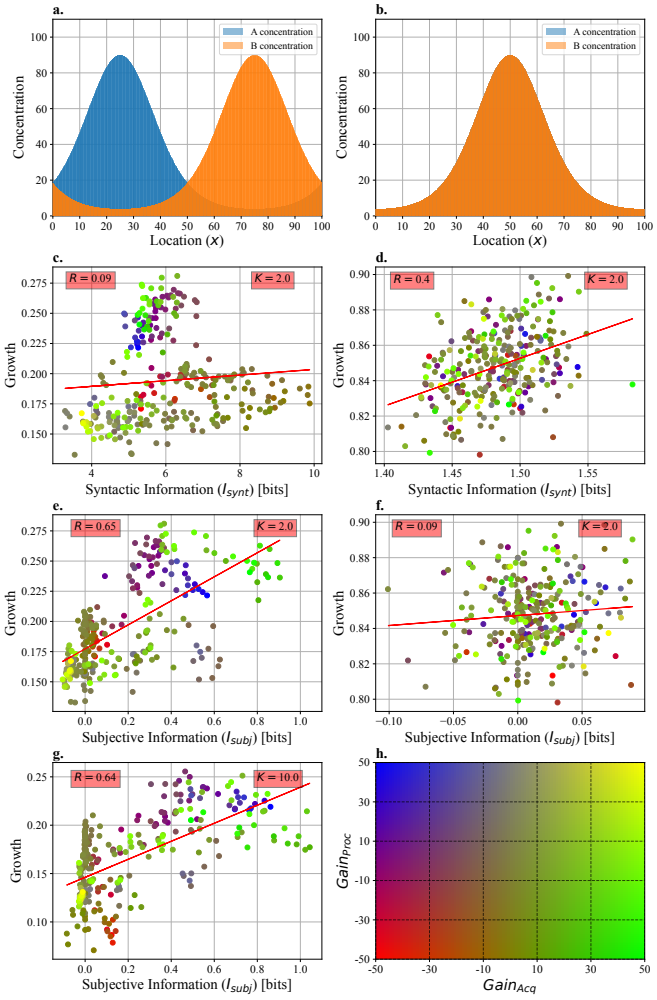

Fig. 3. (a) **Disjoint** von Mises distribution and (b) **identical** von Mises distribution of the molecule concentrations in the environment. (c) Estimated syntactic information with a dissociation coefficient $K = 2.0$ in the disjoint case. (d) Syntactic information for $K = 2.0$ in the identical case. (e) Subjective information for $K = 2.0$ in the disjoint case. (f) Subjective information for $K = 2.0$ in the identical case, where eight negative outliers that were considered to be artifacts of the estimate have been removed from the dataset. (g) Subjective information for the disjoint case when $K = 10.0$. (h) Color-coding legend for figures (c)-(g) of the values for $\mathrm{Gain}_{\mathrm{Proc}}$ and $\mathrm{Gain}_{\mathrm{Acq}}$.

given an environment; rather, it signifies a potential advantage over a system lacking such information.

We propose to characterize a system as non-subjective when the useful information of different environmental signals does not dynamically change across organisms with the same processing strategy, or through their lifetimes. This condition can occur due to a lack of specificity of the useful information with respect to the "needs" (internal state) of the organisms. The system with identical von Mises distributions for the two molecules 'A' and 'B' provides one example. Instead, a system can be said to be subjective when the useful information is strongly specific to the internal state of each organism across sensing strategies. In the latter case, it is advantageous for an

organism to use its internal state to inform the information acquisition and processing strategies. In this case, the emergence of subjective information is observable and positively correlated with a higher growth rate.

VII. Conclusions

In this paper, we propose a novel metric to measure the amount of information in a population of biological individuals (cells) deriving from individual adaptation strategy of their chemoreception, and at the same time useful for achieving growth and survival goals. This type of information is an extension of the concept of useful information, and it was previously found to be an emergent property from the simulation of a biological system with certain characteristics and in specific environmental conditions. In our computational model, each cell is moving, dividing, or dying in an environment from which it requires information. Here, we have extended the computational model to address (i) various possible strategies of the biological system individuals, and (ii) the contribution to the subjective information deriving solely from the strategy in information acquisition, *i.e.*, gradient sensing. We believe this work will contribute to the understanding of the relationship between information and biological goals and provides insights into the different frameworks for information processing in living organisms.

While in this paper we accounted only for the contribution of chemoreception to the subjective information, we believe that other contributions will exist from other biological system components, such as from adaptation strategies in the information processes, as we introduced in our abstract model. Leveraging the notion of subjective information introduced in this paper for the design of organisms that are better adapted to their environments is an interesting target for future work.

References

[1] G. Battail, "Biology Needs Information Theory," *Biosemiotics*, vol. 6, no. 1, pp. 77–103, Apr. 2013.

[2] C. Adami, "Information theory in molecular biology," *Physics of Life Reviews*, vol. 1, no. 1, pp. 3–22, 2004.

[3] A. G. Dimitrov, A. A. Lazar, and J. D. Victor, "Information theory in neuroscience," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 1–5, Feb. 2011.

[4] A. Levchenko and I. Nemenman, "Cellular noise and information transmission," *Current Opinion in Biotechnology*, vol. 28, pp. 156–164, 2014, nanobiotechnology • Systems biology.

[5] C. Harper, M. Pierobon, and M. Magarini, "Estimating information exchange performance of engineered cell-to-cell molecular communications: A computational approach," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 729–737.

[6] A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, "Information theory of molecular communication: directions and challenges," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 2, pp. 120–142, 2016.

[7] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The internet of bio-nano things," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 32–40, 2015.

[8] A. Marcone, M. Pierobon, and M. Magarini, "Parity-check coding based on genetic circuits for engineered molecular communication between biological cells," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6221–6236, 2018.

[9] H. M. Kiah, G. J. Puleo, and O. Milenkovic, "Codes for DNA sequence profiles," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3125–3146, 2016.

[10] F. Ratti, C. Harper, M. Magarini, and M. Pierobon, "Optimizing information transfer through chemical channels in molecular communication," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1–6.

[11] P. J. Thomas, D. J. Spencer, S. K. Hampton, P. Park, and J. P. Zurkus, "The diffusion-limited biochemical signal-relay channel," in *Advances in Neural Information Processing Systems*. Citeseer, 2004, pp. 1263–1270.

[12] J. M. Kimmel, R. M. Salter, and P. J. Thomas, "An information theoretic framework for eukaryotic gradient sensing," in *Advances in neural information processing systems*, 2007, pp. 705–712.

[13] M. Vergassola, E. Villermaux, and B. I. Shraiman, "'infotaxis' as a strategy for searching without gradients," *Nature*, vol. 445, no. 7126, pp. 406–409, 2007.

[14] P. J. Thomas and A. W. Eckford, "Capacity of a simple intercellular signal transduction channel," *IEEE Transactions on information Theory*, vol. 62, no. 12, pp. 7358–7382, 2016.

[15] A. S. Moffett, N. Wallbridge, C. Plummer, and A. W. Eckford, "Fitness value of information with delayed phenotype switching: Optimal performance with imperfect sensing," *Physical Review E*, vol. 102, no. 5, p. 052403, 2020.

[16] J. Hidalgo, J. Grilli, S. Suweis, M. A. Muñoz, J. R. Banavar, and A. Maritan, "Information-based fitness and the emergence of criticality in living systems," *Proceedings of the National Academy of Sciences*, vol. 111, no. 28, pp. 10 095–10 100, 2014.

[17] O. Rivoire and S. Leibler, "The value of information for populations in varying environments," *Journal of Statistical Physics*, vol. 142, no. 6, p. 1124–1166, 2011.

[18] R. Linsker, "Perceptual neural organization: Some approaches based on network models and information theory," *Annual review of neuroscience*, vol. 13, pp. 257–81, 02 1990.

[19] A. Kolchinsky and D. H. Wolpert, "Semantic information, autonomous agency and non-equilibrium statistical physics," *Interface Focus*, vol. 8, 2018.

[20] Y. Bar-Hillel and R. Carnap, "Semantic information," *The British Journal for the Philosophy of Science*, vol. 4, no. 14, pp. 147–157, 1953.

[21] W. Bialek, I. Nemenman, and N. Tishby, "Predictability, complexity, and learning." *Neural Computation*, vol. 13, pp. 2409–2463, 2001, place: US Publisher: MIT Press.

[22] T. S. Barker, M. Pierobon, and P. J. Thomas, "Subjective information and survival in a simulated biological system," *Entropy*, vol. 24, no. 5, 2022.

[23] C. Janetopoulos, L. Ma, P. N. Devreotes, and P. A. Iglesias, "Chemoattractant-induced phosphatidylinositol 3, 4, 5-trisphosphate accumulation is spatially amplified and adapts, independent of the actin cytoskeleton," *Proceedings of the National Academy of Sciences*, vol. 101, no. 24, pp. 8951–8956, 2004.

[24] P. A. Iglesias and P. N. Devreotes, "Navigating through models of chemotaxis," *Current opinion in cell biology*, vol. 20, no. 1, pp. 35–40, 2008.

[25] D. Fuller, W. Chen, M. Adler, A. Groisman, H. Levine, W.-J. Rappel, and W. F. Loomis, "External and internal constraints on eukaryotic chemotaxis," *Proceedings of the National Academy of Sciences*, vol. 107, no. 21, pp. 9656–9659, 2010.

[26] B. Hu, W. Chen, H. Levine, and W.-J. Rappel, "Quantifying information transmission in eukaryotic gradient sensing and chemotactic response," *Journal of statistical physics*, vol. 142, pp. 1167–1186, 2011.

[27] T. S. Barker, "Simulation codes available on github." [Online]. Available: https://github.com/tbarke/BioSimulator

[28] D. Pál, B. Póczos, and C. Szepesvári, "Estimation of rényi entropy and mutual information based on generalized nearest-neighbor graphs," *Advances in Neural Information Processing Systems*, vol. 23, 2010.

[29] G. Doquire and M. Verleysen, "A comparison of multivariate mutual information estimators for feature selection," in *International Conference on Pattern Recognition Applications and Methods*, vol. 2. SCITEPRESS, 2012, pp. 176–185.