Synthetic Cells Extract Semantic Information From Their Environment

Beatrice Ruzzante, *Graduate Student Member, IEEE*, Lorenzo Del Moro,
Maurizio Magarini, *Member, IEEE*, and Pasquale Stano

Abstract—Based on a recently reported operative definition of semantic information, due to Kolchinsky and Wolpert (2018), here we present a research program based on the intersection between synthetic biology, molecular communication, and information theory. In particular, we first describe and emphasize the role of "synthetic cells" as new (bio)technological platform for theoretical and applied investigations; next we simulate a smart drug delivery scenario whereby synthetic cells extract semantic information from their environment (made of cancerous cells, which provide a signal molecule that triggers the production of a cytotoxic drug by the synthetic cell).

Index Terms—Information theory, molecular communication, semantic information, synthetic biology.

I. SEMANTIC INFORMATION AND THE SYNTHETIC BIOLOGY PLATFORM

A. Information and Meaning: From Shannon to MacKay and Beyond

SINCE its introduction in the scientific arena, the concept of "information" has prompted theoretical discussions about its very nature and its use in basic and applied research. Quite probably, one of the reasons comes from the common-sense meaning of the term, which essentially refers to a "communication of knowledge", while its technical meaning – introduced by Shannon in the 1940s – refers to a well-known numerical value, probabilistic in its essence, which can be associated with any message sent by a sender to a receiver. Shannon information has a selective nature, as it associates to a message the probability of being selected from the set of all possible messages (finite in number) from a source. Information thus represents a measure of the decrease of uncertainty for a receiver. Thanks to this definition, information can be measured in a rigorous objective manner.

However, when information is considered according to the Shannon definition only, we discard the primitive meaning of the word information, which actually comes from Latin *in-formare* and it means "giving a form", in the sense of producing a change in the agent that receives information about aspects of reality, not known before. This second facet, pointing to the meaning of messages (i.e., to semantics) is explicitly

absent in the Shannon definition [1], [2]. The first approach to semantic information was made by Carnap and Bar-Hillel in 1952 [3]. Their theory aimed to calculate the amount of semantic information encoded by a sentence in a particular language. The application of such approach is restricted to study the language and unenforceable to physical scenarios. We recently commented on the possible role of MacKay theory of semantic information [4], [5], highlighting those approaches whereby systems (agents) are embedded in *situations*, and their internal states and/or performance, or even their existence, depend on the information exchanged with the environment. In this view, the recent proposal put forward by Kolchinsky and Wolpert (hereinafter, KW) seems particularly relevant to this subject [6]. Here we will briefly sketch a possible application of the KW proposal in synthetic biology, and in particular to bottom-up Synthetic Cells (SCs).

B. The Kolchinsky-Wolpert Approach

In 2018 KW proposed an elegant quantitative measure of semantic information received by any physical system influenced by an external environment [6]. According to the definition given by Rovelli in [7], KW defined semantic information as *the information a physical system has about its environment that is causally necessary for the system to maintain its own existence over time*. As such, semantic information is a subset of *causal necessary* syntactic information that the system needs to preserve its state, or *existence*. Below, we describe the KW model and report the most relevant formulas to evaluate semantic information (see [6] for the complete mathematical description of the problem).

To develop their model, *i)* KW considered a system $\mathcal{X}$ with state space X described by physical variables x, and an environment $\mathcal{Y}$ with state space Y described by physical variables y, subjected to a *stochastic*, *discrete time*, and *first order Markovian* (i.e., the probabilities of the state depends on the previous event) *dynamics* over time. For the sake of simplicity only discrete and finite state spaces were analyzed in [6] even if, as explicitly observed, also continuous state spaces can be considered. In their framework, KW assumed that at the initial time $t = 0$ a correlation exists between the system and the environment, which is described by a joint distribution $p(x_0, y_0)$. This correlation allows to compute the mutual information, measured in bits, between the system and the environment

$$I_p(X_0, Y_0) := \sum_{x_0} \sum_{y_0} p(x_0, y_0) \log_2 \frac{p(x_0, y_0)}{p(x_0)p(y_0)}, \quad (1)$$

where x_0 and y_0 are the physical variables composing the state space X and Y, respectively, and $p(x_0)$ and $p(y_0)$ are the

marginal distributions over the states of the system and of the environment, respectively, at the initial time $t = 0$. In [6] it was further assumed that the system and the environment are subject to coupled (possibly stochastic) dynamics until time t_*, which defines the timescale of interest. Then *ii)*, KW defined the degree of existence of the system as its ability to remain in a low entropy state. In agreement with the principles of statistical physics, to quantify such an ability KW made use of the *viability* (V) function, which is defined as the negative of the Shannon entropy S of the distribution $p(x_t)$ over the states of system $\mathcal{X}$ at time t as

$$V(p_{X_t}) := -S(p_{X_t}) = \sum_{x_t} p(x_t) \log_2 p(x_t). \quad (2)$$

Next *iii)*, KW found a way to scramble the correlations between $\mathcal{X}$ and $\mathcal{Y}$ on the initial configuration ($t = 0$) by introducing a set of *counter-factual partial interventions*. Each *partial intervention* is induced by a particular *coarse-graining function* $\phi(y)$ that specifies the distinctions the system could make about the given fine- (non-coarse-) grained environment [8]. The role of the coarse-graining function is to map each fine-grained state in another possible state. Thus, an intervention modifies the distribution $p(x_0, y_0)$, and consequently also the initial mutual information. The *intervened joint initial distribution* induced by a given coarse-graining function $\phi(y)$ is given by

$$\hat{p}^{\phi}(x_0, y_0) = \hat{p}^{\phi}(x_0|y_0)p(y_0), \quad (3)$$

where

$$\hat{p}^{\phi}(x_0|y_0) := p(x_0|\phi(y_0)) = \frac{\sum_{y_0':\phi(y_0')=\phi(y_0)} p(x_0, y_0')}{\sum_{y_0':\phi(y_0')=\phi(y_0)} p(y_0')} \quad (4)$$

defines the conditional probability of the communication channel through the environment. The sum over y_0' means that the intervened function may map different states in only one and that the system cannot recognize the granularity of the states space of the environment. It is worth observing that the most destructive intervention, called "fully scrambled", leads to have the system and the environment fully uncorrelated and so $\hat{p}^{full}(x_0, y_0) = p(x_0)p(y_0)$. In this particular case the initial mutual information is equal to zero since the system is not able to distinguish the states of the environment. Finally *iv)*, the viability V and the mutual information I_p pairs, estimated for all the intervened distributions, were compared to those of the actual (or non-intervened) distribution. The *optimal intervention* was defined by KW as the one achieving the same value of viability as the non-intervened case at the time of interest t while giving the smallest value of mutual information

$$\hat{p}^{opt}(x_0, y_0) = \arg \min_{\hat{p}^{\phi}:\phi \in \Phi} I_{\hat{p}^{\phi}}(X_0, Y_0)$$
$$\text{s.t} \quad V\left(\hat{p}^{\phi}_{X_t}\right) = V(p_{X_t}), \quad (5)$$

where Φ defines the set of all possible coarse-grained functions. The *stored semantic information* is then defined as the initial mutual information under the optimal intervention

$$S_{stored} := I_{\hat{p}^{opt}}(X_0, Y_0). \quad (6)$$

This operation discriminates the minimum level of distinctions that the system requires about the environment to preserve its existence, thus the causal relation among all the existing relations between the two entities. Of note, the computation of the optimal intervention strictly depends on the time instant t in which the viability is computed.

To recapitulate, the amount of I_p under the optimal intervention was defined by KW as the *stored semantic information* exchanged between the environment and the system since the syntactic information maintained in the optimal intervention is the only needed to preserve the state of existence of the system in the environment. The word *stored* is added by KW to distinguish the semantic information derived from the scrambling of the relation between environment and system at $t = 0$, and computed through the mutual information at $t = 0$, from the *observed semantic information*, which is the semantic information dynamically exchanged between environment and the system, and measured by KW through the use of transfer entropy. From now on, with *semantic information* we refer to the KW's optimal intervention.

C. Bottom-Up Synthetic Cells

Bottom-up SCs can be defined as man-made molecular systems partially resembling biological cells both in structure and function [9]. The bottom-up approach consists in the assembly of complex cell-like systems starting from simple components, such as proteins, nucleic acids, lipids, and sugars, into environmental-responsive structures. The resulting SCs are not biological cells manipulated ad hoc, but bio-chemical systems built from scratch. The bottom-up technology allows the construction of non-trivial SCs, but despite these advancements these systems are not yet alive. They should be best considered as a sort of molecular machines rather than minimal organisms [10]. Several cytomimetic functions have been implemented in SCs: sending and receiving chemical signals [11], producing chemical energy (ATP) [12], allowing nutrients enter through large pores [13], and producing proteins [14].

SCs represent an important new synthetic biology tool with intriguing roles in applied research. One of the most obvious applications refers to their potential use as "smart" agents for drug/protein/gene delivery [15]. In this context, SCs can be described as programmable vehicles that interact with a biological *milieu*. During the "travel" toward the target location, and in its nearby, SCs are supposed to engage chemical communication and perception in order to control their own internal operations – for example the production of a therapeutic compound [16], its release, or even their own self-destruction [15]. Hypothetical SCs capable of behaving in this manner have been scheduled in the synthetic biology agenda, but they are still too complex for the current standards. These limitations, however, do not prevent investigations based on modeling, as the one discussed in this paper. Moreover,

numerical simulations can guide future experimental design by predicting how SCs respond to environmental variables.

D. Plan of the Work

The scenarios we are going to discuss embrace semantic information theories and the potential interest in SCs as smart drug delivery agents. We will describe SCs as platforms to investigate semantic information, with the long-term goal of placing side-by-side simulations and experiments. This study can be considered as one of the first attempts to introduce and to find an application for the concept of semantic information in systems where the transmission of information is based on molecules, i.e., molecular communication-based systems. We present preliminary numerical results referred to a simple but effective SC model inspired by (but not identical to) the KW elegant theory. Convinced that the possible connection between SCs and semantic information is very intriguing, we will explore it deeply in future investigations, while this initial report aims at stimulating discussion in the field and attract attention of experimentalists and theoreticians.

II. Semantic Information in SC Scenarios

We devised a scenario that allows the application of KW approach to an SC characterized by a specific internal behavior. The SC acts as an agent ($\mathcal{X}$) embedded in an environment ($\mathcal{Y}$). SCs can be engineered to complete specific tasks that could be therapeutically exploited, for instance. In the case we considered, SC has the aim to interact with chemicals hypothetically produced by Cancerous Cells (CCs) present in the environment, and to respond by establishing a dependent dynamics that end with the production by SC of *toxin* molecules that kill the CCs (Fig. 1). The most important aspects to be considered for the identification of semantic information in models are *i)* the choice of an agent whose behaviour can be affected by *observing* the environment, *ii)* the establishing of the *dynamics*, and *iii)* the determination of a measurement for the discrimination of meaningful from meaningless information, thus the choice of the KW's *V*. With reference to the very last point, KW relate the concept of viability to the ability of the agent to remain far from the equilibrium state. In this specific case, we consider an SC in an equilibrium state when unable to perform the task for which it is designed. Accordingly, for our purposes it is convenient to define an SC "viable" or "dead", respectively, when it is or it is not able to release toxin molecules to kill CCs. The next subsections describes the SC we devised and simulated. The choice of simulated data is purely arbitrary, however the same model could be generalized.

A. Environment

The environment $\mathcal{Y}$ is described by a single state variable $Y_S = \{0, \ldots, 5\}$, which defines the level of signaling (S) molecules released in the environment by the CCs. The value $Y_S = 0$ corresponds to the absence of signal in the environment, which means that CCs are not present or the signalling molecules have been degraded (see Section II-C).
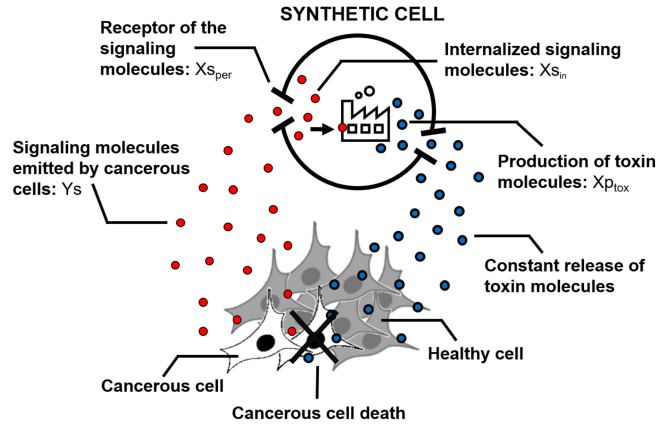

Fig. 1. Graphical representation of the simulated scenario. A *Synthetic Cell* continuously releases toxin molecules, whose production-rate changes according to the amount of *signaling molecules* in the environment.

B. SC Structure

The SC behavior includes four operations: perception of the S levels in the environment; internalization of S; *de novo* production of toxin molecules; and release of toxin molecules. Accordingly, the SC state space $\mathcal{X}$ consists of three separate degrees of freedom, indicated as $X = X_{S_{in}} \times X_{S_{per}} \times X_{ptox}$. Specifically, $X_{S_{per}} = \{0, \ldots, 5\}$ is the variable associated with the perception of S in the outer environment, and biologically represents the "activation" of receptors of S – those that lie on the SC membrane and detect signals from the environment. For simplicity, we assume that $X_{S_{per}}$ is always equal to Y_S, (perfectly responding receptors). $X_{S_{in}} = \{0, \ldots, 5\}$ is the amount of S internalized by SCs that is directly related to $X_{S_{per}}$ as will be described in Section II-C. Finally, $X_{ptox} = \{0, \ldots, 8\}$ is the amount of toxin molecules inside the SC. These molecules are supposed to be membrane permeable so that they are passively released by SCs in the environment (X_{ptox} decreases).

C. SC Dynamics in the Environment

By analogy with traditional drug delivery systems, in our model we suppose that SCs already have a toxin payload at $t = 0$. Moreover, as a smart drug delivery agent, SCs display a behavior that is affected (controlled) by the presence of CCs via their chemical exudates. Specifically, when exposed to CC-secreted S molecules, SCs start a *de novo* production of toxin thanks to internal molecular machineries (X_{ptox} increases). Such internal dynamics increases the possibility that CCs die because of the augmented overall exposure of CCs to the toxin. As mentioned, as far as $X_{ptox} \neq 0$, an SC constantly releases toxins in the environment also in absence of CCs. This last state, related to SC death, brings to a high entropy that is associated with the reduction of system viability (please see Section I-B for the computation of viability). The imposed model dynamics are such that:

1) at $t = 0$ the amount of S in the environment is uniformly distributed over $1, \ldots, 5$ (i.e., S cannot be absent in the system); the SC has perfect information about the amount of S in the environment (i.e., $X_{S_{per}} = Y_S$); $X_{S_{in}} = 0$; $X_{ptox} = 5$;
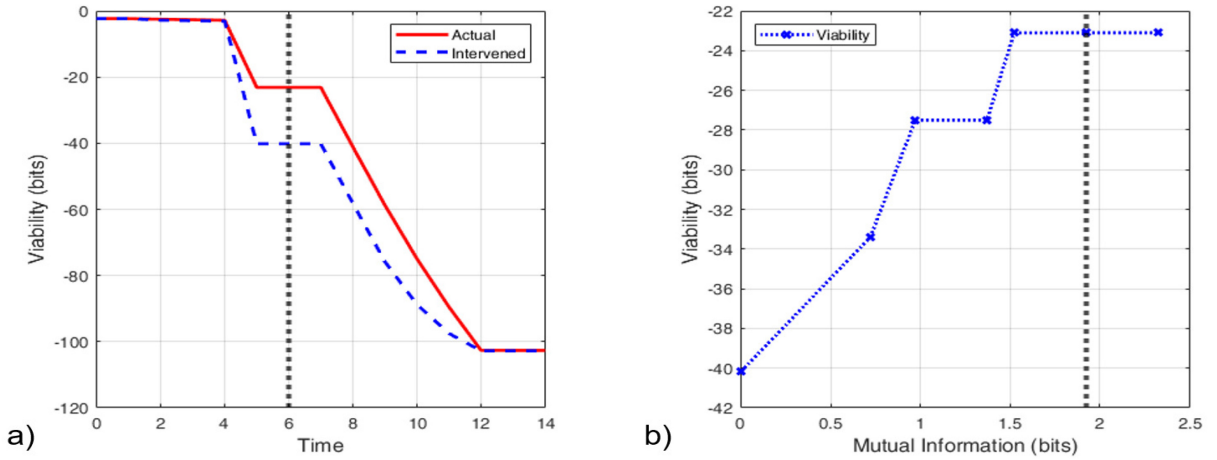
Fig. 2.   Illustration of the KW approach applied to SC. Under the actual distribution, the SC has perfect knowledge of the S level in the environment at $t = 0$. (a) Plot of viability values over time under both the actual and the fully scrambled (i.e., complete uncorrelation between environment and system) intervened distributions. The computed data for the other 31 intervened distributions are not shown for the sake of clarity. The vertical dashed line corresponds to the time of interest ($t = 6$ time steps), when the semantic information is computed. (b) Information vs viability curve for $t = 6$. The vertical dashed line indicates the amount of stored semantic information that is the minimum value of mutual information needed to preserve the viability value of the agent.

2) the level of Y_S remains constant up to t_* unless *i)* S is degraded, *ii)* S is transported away (e.g., by the flux in a blood vessel), or *iii)* the information flow between environment and SC is interrupted as in case of the degradation of the receptors on the SC membrane: these phenomena bring $Y_S = 0$, and can occur with a small fixed probability equal to 0.1 at each step;

3) $X_{S_{in}}$ increases of 1 unit per time step as long as $X_{S_{per}} > X_{S_{in}}$; it corresponds to the internalization of S;

4) the toxin, if present in the SC, is continuously released in time with a fixed rate equal to 1 level per time step;

5) the toxin is produced *de novo* according to the level of S internalized: this operation relates the response of SC to the presence of CCs; when S is present, it leads to an increase of the number of toxin molecules in the SC, and, hence, the number of molecules released in the environment with a fixed rate of 1 level per time step, thus the overall exposure of CCs to toxin;

6) SC produces instantaneously the toxin in an amount equal to Y_S if $Y_S, X_{S_{in}} \leq 1$, hence if enough S molecules have been internalized with respect to what is present in the environment;

7) once the toxin is produced, we assume that the external molecules of S in the environment cannot affect again the SC, so $Y_S = X_{S_{per}} = 0$;

8) when $X_{ptox} = 0$, the SC enter in the above-defined "dead" state because it cannot accomplish its task, and it is attributed a value of internal entropy equal to 100 bits: this corresponds to the KW equilibrium macrostate, with a large entropy [6];

9) simulations were carried out for $t_* = 14$ time steps and the viability in (2) was evaluated at $t = 6$.

Of note, the threshold values, the number of levels for each variable, and the simulation time are purely arbitrary, since this kind of SC still does not exist. However, when experimental data will be available, it would be possible to implement this model to compare the simulation results with the reality.

III. RESULTS

As mentioned, given the internal dynamics of an SC, the KW approach was used to compute the semantic information exchanged between environment and SC. Specifically, a coarse-graining function grouping the possible states of Y_s determined all the possible interventions [6]. Next, for all the intervened and actual distributions, the mutual information I_p and the V function were computed.

Figure 2 shows simulation results for the SC scenario described in Section II. Specifically, Fig. 2a shows the *viability vs time* curve for the actual distribution and for the fully scrambled intervened distribution. The distribution $p(x_t)$ used in (2) to evaluate the viability was obtained by marginalization with respect to x_t of the multivariate joint probability $p(x_t, y_t, \ldots, x_0, y_0) = \prod_{k=1}^{t} p(x_k, y_k | x_{k-1}, y_{k-1})$, which descends from the first order Markovianity of the problem. Each term in the product was computed as $p(x_k, y_k | x_{k-1}, y_{k-1}) = p(x_k | x_{k-1}, y_{k-1}) p(y_k | x_k, x_{k-1}, y_{k-1})$, where $p(x_k | x_{k-1}, y_{k-1})$ is the response of the system to the previous state of itself and of the environment, while $p(y_t | x_t, x_{t-1}, y_{t-1})$ is the response of the environment to the previous state of itself and the system, as well as the current state of the system. The vertical dashed line corresponds to the time instant $t = 6$. Of note, when approaching the analysis of semantic information according to the method of KW, particular attention must be given to the timescale of the analysis because of the time-dependency of V: the timescale must be chosen according to *i)* the experimental data, *ii)* the SC functioning and goal, and *iii)* the expected behavior of the SC to maximize the computation of semantic information content. Even if at the end of the dynamics the viability values for the actual and the intervened distribution reach the same result, given the fact that independently from connections with the environment $X_{ptox} = 0$ at the end of the simulation because of the constant release of the toxin molecules, we can observe a decrease of the viability of the intervened case with respect to the actual case. This finding

is coherent with KW results, and with what was expected by drawing the model: the coarse graining at $t = 0$ decreases the ability of the SC of *interpreting* the information transmitted by the environment (i.e., the presence of the *S*), which can increase the production of the toxin, hence the viability of the SC. The decrease of viability by a coarse-graining is a hint of a premature production of X_{ptox} that leads to a dead macrostate. Although this fact seems to be positive, it is not since we want a toxin production coherent with the presence of *S*. In Fig. 2b, the information vs viability curve shows the value of *semantic information* of the SC in the environment that is $I_S = 1.92$ bits with respect to the total mutual information equal to $I_p = 2.32$ bits. The value corresponds to the mutual information obtained for two optimal interventions, which are given by the coarse graining of *S* levels $\{1, 2\}$ and $\{0, 1, 2\}$, respectively. The presence of two optimal interventions is given by the nature of the level $S = 0$, which is related to its absence in the environment, and impossible at $t = 0$. For this reason, the level $S = 0$ does not carry both mutual information or viability, thus we can consider as optimal intervention the coarse graining of *S* level $\{1, 2\}$. The shown results give some hints about the dynamics of the SC and how it perceives the environmental variables. The SC here described cannot distinguish the environmental levels (or concentrations) of *S* between $\{1, 2\}$ because of the dynamics of the system. In particular, since the SC can produce the toxin if $Y_S - X_{S_{in}} \leq 1$, the two *S* levels of the optimal intervention lead to the production of the toxin at the same time step. This means that not all the information extracted by the SC from the environment is semantic, thus part of it does not affect SC behavior. Other results could be obtained by changing the behavior of the SC, for example by inducing the releasing of toxin in presence of the *S* (data not shown for lack of space).

Although simplistic, this model could be regarded as a first step towards the implementation of a predictive model to assess the behavior of SC in response to variation of environmental parameters. The final aim is to *i)* use realistic data describing the behavior of SC to implement the model, *ii)* predict the adaptation of SC to environmental variables and, hence, *iii)* determine the amount of semantic information exchanged by SC and environment to determine the sensitivity of SC for accomplishing the goal for which it was designed, in this case, killing CCs.

IV. Concluding Remarks

In this letter we have shown how to quantify KW semantic information in an SC-based scenario. We suggest that the development of this model (and similar ones) will further enrich the fields of synthetic biology and molecular communications (and their fecund intersection), fostering future theoretical and experimental developments. In our simple yet informative model, an SC is situated in an environment where it engages chemical communicative dynamics with cancerous cells, and the amount of information that is causally necessary to complete its task has been computed. The basic steps of the procedure (that we intend to optimize in next studies) include the definition of an SC mechanism, an environment in which the SC is situated, a viability function, and the timescale of the simulated dynamics. In particular, we intend to investigate more complex and realistic models, and explore alternative ways to compute SC viability for task-guided systems.

References

[1] R. K. Logan, "What is information?: Why is it relativistic and what is its relationship to materiality, meaning and organization," *Information*, vol. 3, no. 1, pp. 68–91, 2012.

[2] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 210–219, Feb. 2022.

[3] R. Carnap and Y. Bar-Hillel, "An outline of a theory of semantic information," Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Rep. 247, Oct. 1952.

[4] D. M. MacKay, *Information, Mechanism and Meaning*. Cambridge MA, USA: MIT Press, 1969.

[5] M. Magarini and P. Stano, "Synthetic cells engaged in molecular communication: An opportunity for modelling Shannon- and semantic-information in the chemical domain," *Front. Commun. Netw.*, vol. 2, Sep. 2021, Art. no. 724597.

[6] A. Kolchinsky and D. H. Wolpert, "Semantic information, autonomous agency and non-equilibrium statistical physics," *Interface Focus*, vol. 8, no. 6, 2018, Art. no. 20180041.

[7] C. Rovelli, "Meaning and Intentionality = Information + Evolution," in *Wandering Towards a Goal: How Can Mindless Mathematical Laws Give Rise to Aims and Intention?* (The Frontiers Collection), A. Aguirre, B. Foster, and Z. Merali, Eds. Cham, Switzerland: Springer, 2018, pp. 17–27.

[8] J. Rauh, P. K. Banerjee, E. Olbrich, J. Jost, N. Bertschinger, and D. Wolpert, "Coarse-graining and the blackwell order," *Entropy*, vol. 19, no. 10, p. 527, 2017.

[9] P. L. Luisi, "Toward the engineering of minimal living cells," *Anat. Rec.*, vol. 268, no. 3, pp. 208–214, 2002.

[10] P. Stano, "Exploring information and communication theories for synthetic cell research," *Front. Bioeng. Biotech.*, vol. 10, Jul. 2022, Art. no. 927156.

[11] R. Lentini et al., "Two-way chemical communication between artificial and natural cells," *ACS Central Sci.*, vol. 3, no. 2, pp. 117–123, 2017.

[12] E. Altamura et al., "Chromatophores efficiently promote light-driven ATP synthesis and DNA transcription inside hybrid multicompartment artificial cells," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 7, 2021, Art. no. e2012170118.

[13] V. Noireaux and A. Libchaber, "A vesicle bioreactor as a step toward an artificial cell assembly," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 51, pp. 17669–17674, 2004.

[14] P. Stano, "Gene expression inside liposomes: From early studies to current protocols," *Chemistry*, vol. 25, no. 33, pp. 7798–7814, 2019.

[15] P. R. Leduc et al., "Towards an in vivo biologically inspired nanofactory," *Nat. Nanotechnol.*, vol. 2, no. 1, pp. 3–7, 2007.

[16] N. Krinsky et al., "Synthetic cells synthesize therapeutic proteins inside tumors," *Adv. Healthc. Mater.*, vol. 7, no. 9, 2018, Art. no. e1701163.