Inside looking out

Fernando Rodriguez

University of Sussex
fr97@sussex.ac.uk

Abstract

One of the defining, foundational axes of enactivism was its emphasis on the necessary relation between cognition and phenomenological experience, assumedly rooted on the particular, organizationally recursive nature of autonomous systems. However, in spite of many advances, there is no conclusive understanding about the emergence of this experiential dimension yet; a conundrum that has lead to contrasting positions within the framework. In this context, we suggest that an enactive, not fully committed interpretation of ideas from the Integrated Information Theory of Consciousness (IIT) may result fruitful; In particular, the formal notions of intrinsic information and integration as indicative of an intrinsic perspective and emergence respectively.

Autonomy and phenomenology

Maturana and Varela (1973) claim that, given their recursive nature, autopoietic systems possess an intrinsic *identity* underlied by the autonomous subordination of structural changes to the preservation of their organization; determining an independent and self-contained (biological) phenomenological subdomain (Maturana and Varela, 1973, p.69,p.110). In later work, and in order to account for a general notion of *autonomy* beyond cellular specificities, Varela (1979) introduced the formal concept of *organizational closure*, characterizing autonomous behavior in terms of operations within a self-referential space of transformations.

In *The Embodied Mind* (Varela et al., 1991) autonomy is characterized in enactive terms, in light of an embodied and embedded view on cognition. This is illustrated with Bittorio, a minimally autonomous cellular automata described as capable of *bringing forth a domain of significance* by selectively enacting external regularities (hence, displaying cognitive behavior), but considered obviously devoid of experience (e.g. in contrast to color perception) (Varela et al., 1991, p.150-157). This gap is the crux of the matter; why/how some varieties of autonomy would entail phenomenological experience and not others? While Varela (1997) claims that is the *autonomy of living systems* that provides a referential perspective for meaning and intentionality in phenomenological terms. And that experience, as for us, results from the particular nested and sensorimotor nature of the autonomy of the nervous system; there is still no decisive justification for this assumption. This has lead to, very broadly speaking, *traditional* positions (Di Paolo, 2005; Egbert and Barandiaran, 2014; Barandiaran, 2017; Di Paolo et al., 2017; Froese and Taguchi, 2019), more or less *radical* positions (Hutto and Myin, 2012, 2017; Abramova and Villalobos, 2015; Villalobos and Silverman, 2018), and other approaches leaning towards more representational explanatory stances (Clark, 2016; Seth and Tsakiris, 2018).

Formalizing autonomy

For an autonomous system, its dynamic interactional selectivity will determine different state transitions, mediated by environmental circumstances, so that $(st_i, e_k) \mapsto st_x$ (where st_i and st_x are states of the system, and e_k is an environmental state). Because every system is influenced by environmental interactions, such mappings are normally many-to-one; therefore, whenever an enaction takes place, there is a categorization determined by the particular nature of the autonomy, which implicitly specifies a certain structural instantiation, with some inherent sensitivity and possibilities for action; It is in this sense that an enaction is simultaneously a distinction and an action. For a deterministic case, we can define an *enaction set* as the set of all environments enacted in the same way, a related function that maps pairs of states to these sets, and sets of *enaction categories* containing the valid transitions available to a system:

$$es(st_i, st_x) = \{e_k : (st_i, e_k) \rightarrow st_x\} \tag{1}$$

$$f_{es}(st_i, st_x) := \{e_k : (st_i, e_k) \rightarrow st_x\}. \tag{2}$$

$$ec(st_i, st_x) := (st_i, st_x, f_{ec}(st_i, st_x)). \tag{3}$$

In spite of their simplicity, these definitions allow us to determine relevant properties of a (relatively simple) autonomous system, such as its organization, selectivity, or structural degeneracy. An exhaustive (and wonderful) illustration of the dynamic properties of autonomy, in the context of autonomous patterns in the Game of Life, can be found in (Beer, 2004, 2014, 2020a,b).

Integrated information?

Even if powerful, our current formal descriptions of autonomous systems seem unable to capture some important qualities associated to a phenomenological dimension; in particular, the above mentioned notions of phenomenological (operational) domain and the emergence of a global coherence. In this context, to explore the possibility of incorporating concepts from the Integrated Information Theory of Consciousness (IIT), such as *intrinsic information* and *integration*, probably by making an enactive reinterpretation, may result fruitful.

The IIT considers selectivity to be a requirement for *intrinsic information*, as it underlies the capacity of a mechanism to constraint past and future system's states: *its cause-effect power*. Cause-effect information is conceived as a measure of how relevant a change is, from the *perspective of the system itself* (Oizumi et al., 2014) which is given by the minimum (shared) between cause (ci) and effect (ei) information. For simplicity's sake, we will make use of the example system presented in Oizumi et al. (2014) (fig. 1), where, given a system of mechanisms A,B and C; cause and effect information can be obtained from:

$$ci = D(p(ABC^p \mid A^c = 1) \mid\mid p^{uc}(ABC^p)) \quad (4)$$

$$ei = D(p(ABC^f \mid A^c = 1) \mid\mid p^{uc}(ABC^f)) \quad (5)$$

Here, $p(ABC^p \mid A^c = 1)$ and $p(ABC^f \mid A^c = 1)$, called cause and effect repertoires, are the constrained probability distributions for past and future system's states, given that $A = 1$; the superscripts ^{p},^{c} and ^{f} stand for past, current and future, and ^{uc} denotes unconstrained distributions; whereas D, represents the distance between distributions, measured using the earth mover's distance (EMD) method.

In turn, integration implies that a system able to support emergent experience must be an irreducible whole, not far from the enactive premise of an emergent identity, even if the specific sense of emergence (Chalmers, 2011) is not always unequivocal. Unfortunately, as described by Mediano et al. (2018), different methods for measuring the integration of a system yield inconsistent results, which is obviously problematic for conclusive interpretations. Although less specific, we could consider using a more reliable measure for integration like that from De Rosas et al. (2020):

$$\Psi_{t,t'}(V) := I(V_t; V_{t'}) - \sum_j I(X_t^j; V_{t'}) \quad (6)$$

While both approaches consider systems to exist intrinsically and to be intrinsically determined (even if susceptible to external perturbations) by means of their organization (causal structure); practically, this can be implemented in different manners. Indeed, IIT's repertoires are evaluated by fixing the environment and defining a new transition matrix whenever is needed; this however, if we conceive system's transitions in terms of enaction, would be obscuring the fact that such transitions entail (syntactic, but phenomenologically relevant) distinctions. More concretely, if, for instance, we change the environment (element D=0/1) from figure 1, we could certainly specify two transition matrices, but isolated these are unable to reflect the full range of available enactions, which are closer to be the actual distinctions from the perspective of the system (see table 1).

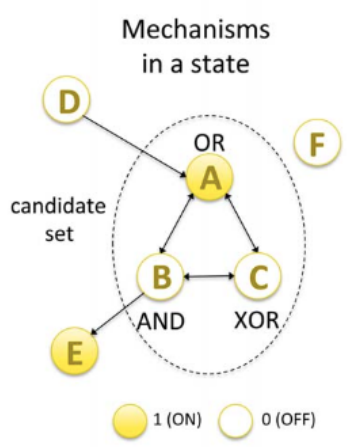

Figure 1: Minimal example system introduced by Oizumi et al. (2014) in the last (3.0) version of IIT.

ec	st_i	env_k	st_x
e1	000	$\{env_1\}$	000
e2	000	$\{env_2\}$	100
e3	001	$\{env_1, env_2\}$	100
e4	010	$\{env_1, env_2\}$	101
e5	011	$\{env_1, env_2\}$	101
e6	100	$\{env_1\}$	001
e7	100	$\{env_1\}$	101
e8	101	$\{env_1, env_2\}$	111
e9	110	$\{env_1, env_2\}$	100
e10	111	$\{env_1, env_2\}$	110

Table 1: The superset of enaction categories available to the system ABC after taking both environmental conditions (ABCD, for D=0 and D=1) into consideration

Thus, we could define a slightly different cause and effect repertoires as distributions of the enaction categories that have led to-, or will inform of affordances with respect to the current state of a mechanism. Of course, in order to ensure congruence when quantifying information, the unconstrained distributions must change accordingly.

References

Abramova, K. and Villalobos, M. (2015). The apparent (ur-)intentionality of living beings and the game of content. *Philosophia*, 43:651–668.

Barandiaran, X. (2017). Autonomy and enactivism: Towards a theory of sensorimotor autonomous agency. *Topoi*, (36):409–430.

Beer, R. (2004). Autopoiesis and cognition in the game of life. *Artificial Life*, (10):309–326.

Beer, R. (2014). The cognitive domain of glider in the game of life. *Artificial Life*, 20:183–206.

Beer, R. (2020a). Bittorio revisited: Structural coupling in the game of life. *Adaptive Behavior*, 28(4):197–212.

Beer, R. (2020b). An integrated perspective on the constitutive and interactive dimensions of autonomy. *Proceedings of the ALIFE 2020: The 2020 Conference on Artificial Life*, July 13-18:202–209.

Chalmers, D. (2011). Strong and weak emergence. *In: The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion. Oxford University Press.*, pages 244–256.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action and the Embodied Mind*. Oxford University Press.

De Rosas, F., Mediano, P., Jensen, H., Seth, A., Barret, A., and Carthart-Harris, R. (2020). Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. *PLOS. Computational Biology*, 16(12).

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, (4):429–452.

Di Paolo, E., Burhmann, T., and Barandarian, X. (2017). *Sensorimotor Life: An enactive proposal*. Oxford University Press.

Egbert, M. and Barandiaran, X. (2014). Modeling habits as self-sustaining patterns of sensorimotor behavior. *Frontiers in Human Neuroscience*, 8(590):1–15.

Froese, T. and Taguchi, S. (2019). The problem of meaning in ai and robotics: Still with us after all these years. *Philosophies*, 4(2).

Hutto, D. and Myin, E. (2012). *Radicalizing Enactivism. Basic minds without content.* MIT Press.

Hutto, D. and Myin, E. (2017). *Evolving Enactivism. Basic Minds Meet Content*. MIT Press.

Maturana, H. and Varela, F. (1973). *Autopoiesis: the organization of the living. [De maquinas y seres vivos. Autopoiesis: la organizacion de lo vivo]. 7th edition from 1994.* Editorial Universitaria.

Mediano, P., Seth, A., and Barret, A. (2018). Measuring integrated information: Comparison of candidate measures in theory and simulation. *Entropy*, 21(1):17.

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLOS Computational Biology*, 10(5).

Seth, A. and Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences*, 22(11):969–981.

Varela, F. (1979). *Principles of Biological Autonomy*. North Holland.

Varela, F. (1997). Patterns of life: Intertwining identity and cognition. *Brain cognition*, (34):72–87.

Varela, F., Rosch, E., and Thompson, E. (1991). *The embodied mind: Cognitive science and human experience*. The MIT Press.

Villalobos, M. and Silverman, D. (2018). Extended functionalism, radical enactivism and the autopoietic theory of cognition: prospects for a full revolution in cognitive science. *Phenomenology and the Cognitive Sciences*, 17:719–739.