

The Apparent (Ur-)Intentionality of Living Beings and the Game of Content

Katerina Abramova^{3,4} · Mario Villalobos^{1,2}

Received: 20 June 2014 / Revised: 1 November 2014 / Accepted: 21 May 2015 /

Published online: 20 November 2015

© Springer Science+Business Media Dordrecht 2015

Abstract Hutto and Satne, *Philosophia* (2014) propose to redefine the problem of naturalizing semantic content as searching for the origin of content instead of attempting to reduce it to some natural phenomenon. The search is to proceed within the framework of Relaxed Naturalism and under the banner of teleosemiotics which places Ur-intentionality at the source of content. We support the proposed redefinition of the problem but object to the proposed solution. In particular, we call for adherence to Strict Naturalism and replace teleosemiotics with autopoietic theory of living beings. Our argument for these adjustments stems from our analysis of the flagship properties of Ur-intentionality: specificity and directedness. We attempt to show that the first property is not unique to living systems and therefore poses a problem of where to place a demarcation line for the origin of content. We then argue that the second property is a feature ascribed to living systems, not their intrinsic part and therefore does not form a good foundation for the game of naturalizing content. In conclusion we suggest that autopoietic theory can not only provide a competitive explanation of the basic responding of pre-contentful organisms but also clarify why Ur-intentionality is attributed to them in such an intuitive manner.

Keywords Mental content · Intentionality · Enactivism · Complex systems

✉ Mario Villalobos
mario.kirmayr@gmail.com

Katerina Abramova
e.abramova@ftr.ru.nl

¹ School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK

² Escuela de Psicología y Filosofía, Universidad de Tarapacá, Arica, Chile

³ Department of Philosophy, Theology and Religious Studies, Radboud University, Nijmegen, The Netherlands

⁴ Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

Introduction

In a recent renewed attack on the Hard Problem of Content (Hutto and Myin 2013), Hutto and Satne (2014) review the state of the play and attempt to redefine the game in order to up our odds of winning. We are invited to consider the possibility that instead of “providing reductive explanations that show how content just *equates to* some decidedly natural phenomenon”, we should be trying to explain the natural *origins* of content. We are also told that the new game is to be played under the banner of Relaxed Naturalism which provides room on the field for empirical insights from a variety of sciences and enables all the players of such an extended team to contribute their unique part to the overall solution. In short, the first basemen are to illuminate the nature of basic contentless intentionality, the second basemen should try to clarify the practices of ascribing content and third basemen essentially are tasked with telling us just how we get from one to another, i.e., how content-involving intentionality that second basemen investigate emerged on the basis of contentless intentionality that first basemen provide us with.

We are broadly sympathetic to Hutto and Satne’s redefinition of the game (henceforth H&S). We do believe that seeking origins of content is a more solvable task than attempting to reduce it. We also believe in the team spirit and the relevance of empirical findings from a variety of natural and social sciences. However, we do not think this should lead to relaxing the naturalistic attitude and we worry that, perhaps in disanalogy to baseball, if the first base is too relaxed, the whole game is likely to fail.

In what follows, we zoom in on the first base as portrayed by H&S and investigate whether it is strong enough to push us onto the winning track. We suggest that in fact it is not, largely because of a conflation of resources that first and other bases bring to the table. Instead of a Relaxed Naturalism, we call for a Strict Naturalism, and instead of teleosemantics, we offer an autopoietic conception of living beings as non-equilibrium multi-stable systems.

The Content Gameplay

The H&S Strategy

Teleosemantics, which H&S take as a starting point for their agenda, is roughly a view that holds that the content of any mental representation is determined by the function that this representation has been selected to perform in the history of an organism using that representation (Millikan 2004). Teleosemantics, proposed as a new first-base player, is teleosemantics with semantic contents replaced by the notion of information-as-covariance. That is, the basic form of intentionality exhibited by organisms, the Ur-intentionality, is nothing more than adaptively responding to features in the environment. Teleosemantic ambitions to account for properly semantic features like truth, reference and intensionality (representing under a description) are rejected as inadequate because selectionist explanations are extensional in nature. What is kept, however, is the idea that “evolved structures can have a kind of ‘specificity’ or ‘directedness’” (p. 21) and that this can account for intentional directedness of organisms toward specific features of the environment.

Thus, there are two elements to the H&S's first base: one is a theory of adaptive responding to the environment based on sensitivity to natural signs (hence *teleosemiotics*), another is an idea that specificity and directedness can be explained through notions of biological functions and selective pressures that shape the organism's behavioral profile (*teleo-semiotics*). Let us examine both of these components.

Natural sign of X, as defined by Millikan (2004) is some Y from which you can learn of that X because there is a trackable connection between them, e.g., if one sees smoke, one can learn that there is fire. Responding to such natural signs, according to Hutto (2008) is a process whereby features of the environment are detected and this enables the production of a successful action for a given organism with given needs. The features in question need to be 'relatively stable' and 're-identifiable' but once the proper machinery is in place, responding can proceed without content-involving conceptual understanding.

Importantly, we are encouraged to accept that such responding does not involve content because there is nothing that is being 'picked up' and processed from the environment. It is merely that the organism is set up in such a way as to enable natural signs to trigger particular responses. The relation between whatever the inner state of the organism is when such process takes place and the feature of the environment, is an objective relation of covariance, not an interpreted relation of representing that the world is thus and so.

For example, a frog that flicks its tongue to a small moving black object (natural sign) does so because such a stimulus reliably enough correlates with flies and allows the frog to ingest them and keep on living. If we wish to say that the inner state of the frog carries information about the presence of the fly, it would only mean as much as to say that the smoke carries information about fire or the number of rings in a tree about the tree's age and vice versa.

Hutto (2008) emphasizes that information-as-covariance is a symmetrical relation: "if the first state of affairs 'carries information' about the second then equally the second 'carries information' about the first". This applies to the tree rings as well as internal frog states. The retreat to the scientifically acceptable notion of statistical rather than content-involving information is meant to lay a lean foundation for contentless intentionality and rid the field of implausible accounts of representational processing in basic organisms. At the same time, however, there is a need to ground the asymmetry of intentionality and that is where teleology seems to come in.

Now, the said teleology appears in Hutto's account in two places. First, there is the teleological workhorse adopted from teleosemantics: "historical facts about what ancestral organisms interacted with in their environments" (Hutto and Satne 2014, p. 23) and how these interactions were beneficial to them (i.e., allowed the organisms to survive and reproduce) can explain current response tendencies. The modification to Millikan's view is that the only truths or mis-representations that are thought to follow here are extensional in nature, intensional errors of judgment must await further development of ascriptional practices.

Second, perhaps more important, place for teleology is in intentional attitudes that have been already shaped by evolution and are operational in the organism that we are seeking to explain. We are told that

Organisms are informationally *sensitive to* and selectively *end-directed at* certain worldly features, objects and states of affairs in a way that explain their success and failure on certain tasks. The intentionality they exhibit is an attitude of the whole organism expressed in their behaviour; it is neither a property of the signs themselves nor of organismic inner states (Hutto 2008, p. 57, emphasis added).

Thus, the reactions are purposeful and directed to certain ends but it is the activity of the whole organism that targets a particular aspect of the world and secures the desired end. The inner states that take part in this process are not contentful though they are world-directed (Hutto 2011).

The move, then, from covariance to asymmetric directedness rests on the fact that while tree rings merely correlate with tree's age, internal states of the frog's eye or brain not only correlate with moving black dots (and therefore flies) but they are used in the process of goal-directed adaptive behavior of the frog as an organism. This allows us to speak of *Ur-intentionality* as constituting the first base.

The Preview of Our Critique

Hutto (2008) powerfully argues that attributing propositional attitudes and proper semantic contents to non-verbal animals is a case of anthropomorphic over-interpretation. In order to counteract this over-interpretation he replaces intelligent behavior based on propositional attitudes with information-sensitive behavior based on (contentless) intentional attitudes. The target article is a plea to adopt the resulting teleosemiotic framework as a foundation for explaining the origin of content - a task made easier by the fact that intentionality is already explained in non-content-involving terms.

Now, our controversial claim here is this: intentionality, even if contentless, is similarly a case of anthropomorphic projection that cannot explain the origin of semantic content, at least not within a framework of Strict Naturalism. To make our claim more precise, we focus on the flagship properties of *Ur-intentionality* that are supposed to characterize the nature of contentless behavior. We consider these properties to be 'specificity' and 'directedness' as evident in this quote from the target article:

Biological explanations can tell us what ancestors of a particular sort of device in fact did *target* and thus what fixed the *range of things* descendant devices now respond to, extensionally speaking. Thus biology provides adequate tools for making sense of something more modest than content – it provides what is needed to understand and explain responses exhibiting a kind of *Ur-intentionality* that results from *the targeted directedness* of past organisms (p. 20, emphasis added).

In what follows we elaborate on both of these features and show that from the standpoint of Strict Naturalism the first is not unique to living systems and therefore poses a problem of where to place a demarcation line for the origin of content (section 3), while the second is a misattribution of properties that are not part of the system and therefore does not form a good first base for the game of naturalizing content (section 4). Additionally, we claim that autopoietic theory can provide a competitive

explanation of the basic responding of pre-verbal organisms. Unfortunately, due to the limits of space we can only touch on that part of our strategy (section 5).

Specificity, Selectivity and the Elusive Origin of Intentionality

Specificity seems to be the tendency of the organisms to respond to certain things in the environment but not others. The emphasis on such selective responding is not new in the context of modern versions of enactivism – an embodied cognitive science paradigm primarily concerned with the origins of the mental (reaching down to life itself). Thompson (2007) explains how meaning can emerge from organism-dependent activity in an inherently meaningless world. For example, a bacterium that requires sugar molecules to stay alive moves in the environment in a way that is sensitive to the presence of these molecules. Every bacterial interaction with the environment is said to be carried out from the perspective of the bacterium's imperative to continue living. As a result, on the side of the environment, sucrose molecules turn from mere physico-chemical elements into something like 'food'. On the side of the organism, the ability to ingest sucrose and not, say, sodium chloride carves out a meaningful portion of the world that can be targeted and responded to. In this 'bringing forth' a meaningful world, enactivism seeks the beginnings of intentionality, meaning and values.

Hutto (2011) sees Thompson's perspective as advantageous in his own agenda of looking for the origins of content because it solves the logical problem faced by teleosemantics. Namely, teleosemantics "requires content to pre-date selection" of the consuming response by the user of natural signs but "requires selection to pre-date content" which does not exist without organisms doing the consumption (Cummins et al. 2006). The enactive solution to this problem is to say that content appears together with the activity of the systems and stress that it is being *created* by them, not *consumed*.

Perhaps H&S would not go as far as Thompson to talk about 'meaning' at the level of bacteria, but surely they would take the example as a clear illustration of contentless intentionality.

It should be apparent that the mere fact of some entity being capable of coming into interaction with just a subset of things that exist in the world cannot be what enactivists take to be the defining characteristic of living organisms which justifies the claim about original intentionality. After all, everything that exists can only interact with certain other things but not others. We know that water mixes with wine and not with oil, but we do not treat that selectivity as an index of Ur-intentionality in water. We distinguish a special class of gases – noble gases – based on the observation that they are highly unreactive. If they do come into reaction, however, to form noble gas compounds, we do not say that whatever they interact with becomes meaningful. Or at least H&S, we guess, would not admit water and gases as Ur-intentional systems.

Enactivists would object that talking about water and gasses is completely beside the point because what is required for an intentional kind of selectivity is that it is subjected to the system's integrity, to its conservation as such. A volume of water mixes with wine and becomes diluted wine. Gases combine and become a new compound which is, and behaves, different from its constituent gases. These entities selectively interact with their environments but in doing so they not only undergo certain changes but

rather lose their identities and become something different, at least at the macro level. The right kind of selective interaction with the environment then is that in which the system, despite undergoing changes, conserves its identity and maintains its adaptation. In interacting with its environment, enactivists would remark, the bacterium is selective precisely because only certain interactions prove to be adaptive and allow its conservation as a bacterium.

But surely this is not yet the whole story, as it is easy to find cases of non-living systems that conserve their identity through changes triggered by certain specific interactions with the environment. For example, a metal bowl left in the presence of air moisture may undergo electrochemical oxidation where its iron molecules will (selectively) react with molecules of water and oxygen to produce iron oxides commonly known as rust. While the bowl retains its form, we may still call it a bowl in spite of these structural changes; i.e., the bowl conserves its identity through its selective interaction with the environment. Enactivists might now ask: is the bowl oxidation adaptive? Can we recognise any adaptation in this process? The reply, as it often is in philosophy, will depend on what one means by ‘adaptation’.

In Maturana’s autopoietic theory, which we favor, the identity of a system is essentially given by its organization (its form or configuration), not by its structure or concrete composition (Maturana 1981). That is, the structure of a system may change without necessarily affecting its identity as a system (i.e., without changing its organization). In this view, adaptation is any process in which a system conserves its identity through structural changes in interaction with its environment. Adaptation, according to this conceptualization, is not an exclusively biological category:

If the organization of a composite unity remains invariant while it undergoes structural changes [...] through its recurrent interactions in its medium [, we say that] its adaptation is conserved [...] Defined in this manner, [...] conservation of adaptation is not peculiar to living systems. It is a phenomenon that takes place whenever a plastic composite unity undergoes recurrent interactions with structural change but without loss of organization (Maturana 1980, pp. xx-xxi).

From this point of view, the oxidizing bowl is clearly an adaptive system in selective interaction with the environment. Is the bowl, then, an Ur-intentional system?

The bowl, it might be pointed out at this point, does not ‘do anything’ to retain its identity, it just ‘sits there’, passively undergoing, better said ‘suffering’, the change. If we agree, it would mean that at the root of intentionality lies not just preservation of identity and selective passive interaction with the environment but rather preservation coupled to some internally generated dynamic, selectively tuned to the world. The reason for this restriction could be that passive physical objects like metal bowls are not able to resist the disintegration, that is, they obey the second law of thermodynamics which, according to the classical metaphor, states that things tend to disorder over time. Ur-intentionality, therefore, would start with systems that are able to locally resist the law through some kind of endogenous dynamics. In complexity science, and after Prigogine’s works (Nicolis and Prigogine 1977; Prigogine 1980), such systems are usually called dissipative structures.

A dissipative structure, in simple terms, is a system whose dynamic organization exhibits stability in far-from-thermodynamic equilibrium conditions (this point will be retaken in the next section with respect to directedness). Examples of dissipative structures, or dynamic patterns, as Kelso (1995) prefers to call them, are Benard cells, tornadoes, whirlpools, stars, and flames (Ji 2012; Ulanowicz and Hannon 1987). Dissipative structures are thermodynamically open or semi-open systems; that is, they conserve their organization through (and thanks to) a constant exchange of matter and energy with the environment. Different dissipative structures exhibit different kinds of internal dynamics. Some of them exhibit internal chemical reactions (e.g., oscillatory chemical clocks), some of them purely thermal or mechanic patterns (e.g., convection patterns).

Benard cells, for instance, are convection patterns which emerge in fluids heated from below. When the temperature of the bottom and top layer of water is the same, the liquid is uniform and at a stable equilibrium (any small perturbation to its state is transient). Once the temperature of the bottom layer is increased, the process of conduction is established. The small differences in temperature are damped by the system and it remains stable. With further increase of temperature, however, the convection process takes off, in which the warmer portion of fluid expands and rises to the top while the colder top portion falls down. This microscopic movement organizes spontaneously into macroscopic structures called convection cells, which keep the system in a stable state in far-from-thermodynamic equilibrium conditions. The fluid as a system self-organises in Benard cells which maintain their identity through a constant heat flux (unless the temperature is increased even further at which point they disintegrate).

One could object that the heat is given to convection cells from outside, they do not produce order out of their own resources (Ji 2012). There are, however, other natural phenomena that could be called self-maintaining dissipative systems. Juarrero (2002), for example, describes autocatalytic reactions in which the product of a reaction is used to activate that same reaction which can lead to runaway effects and a new mode of organization. The organization emerges out of own internal dynamics and creates a system that is selective in its reactions and stable against environmental perturbations. Typical examples of such autocatalytic systems would be flames (Bickhard 2000; Ji 2012; Saborido et al. 2011), and thermonuclear reactions in stars (Ulanowicz and Hannon 1987). Interestingly, most of the authors mentioned here believe they see something quite special in these kinds of dissipative systems. They suggest that in such autocatalytic systems we might find the natural origins of purposiveness (Juarrero 2002), a kind of minimal (proto) teleology (Bickhard 2000; Saborido et al. 2011) or telos (Ulanowicz 2009).

This point is close to the agenda of enactivists, as well as the strategy of Hutto and Satne (2014), which we discuss here. At least, we come very close to the selectivity and telos of teleosemiotics:

As [dissipative autocatalytic processes] *select* for inclusion in the web molecules that *enhance overall activity*, autocatalytic cycles ‘aim’ at greater performance by constantly pruning and streamlining their pathway structure. [...] a precursor of teleology is detectable in the way such structures of process operate (Juarrero 2002, p. 127, emphasis added).

We might disagree with the particular conclusion of Juarrero. However, given the considerations above, if our view on the origins of content depends on Ur-intentionality, we need to either accept that Ur-intentionality starts with dissipative autocatalytic cycles (i.e., admit that candle flames and stars are Ur-intentional) or specify *why* this is not the case. Why, for example, Ur-intentionality needs to start with living, behaving organisms. Otherwise, it is not clear what the boundaries of the first base are and whether the choice is a principled one or arbitrary. The latter option would not be good news for anyone hoping to ground content in some basic form of intentionality.

The point about demarcation lines is made even more relevant if we recognize that living beings actually belong to the group of autocatalytic dissipative systems discussed by Juarrero and other complexity folk (Prigogine and Stengers 1984). Living beings are semi-open dissipative structures in constant exchange of matter and energy with the environment (Bechtel and Richardson 2010), whose internal dynamics consists in a rich chemical network of autocatalytic processes (Prigogine and Stengers 1984).

Wishing to stick to their guns, H&S could respond that perhaps it is not merely selectivity that does the trick but directedness, the second feature we distinguished within the concept of Ur-intentionality, and so it is to that feature that we now turn.

The Reality and Appearance of Directedness

Varieties of Directedness, Appearance and Reality

To recall, in the preceding section we have considered Ur-intentionality to comprise conservation of the system's identity through the selective interaction with the environment. We have shown, after considering several alternatives, that mere selectivity at the level of interactions does not seem to be enough for grounding Ur-intentionality in a non-ambiguous way. Enactivists might reply we have overlooked so far a very important element of Ur-intentionality, namely directedness.

To clarify our discussion, we would like to specify directedness as hiding two further features commonly associated with living organisms. One, which we might call teleological directedness, is closely related to specificity and conservation discussed above. It is directedness in the sense of aiming at a particular state of affairs in the future and behaving so as that state is attained. The second feature has to do with the alleged asymmetrical relationship held between the organism and the environment. It is the organism that is said to be directed, both in its sensory and motor activity, toward the environment and not the other way round. This kind of directedness we might call sensorimotor directedness.

In this section we object to both ways of applying directedness to living beings but for slightly different reasons. In the case of teleological directedness, we argue that the goal-oriented character of living beings is an appearance that results from their high complexity as non-equilibrium multi-stable systems. In the case of sensorimotor directedness, we argue that the world-oriented character of living beings is an appearance that results from the particular point of reference adopted by the observer in her descriptions, who overlooks the closed and directionless functional organization of the sensorimotor system. In both cases, we conclude, the alleged directedness is a feature

ascribed by the observer, i.e., a second-base play, not an intrinsic and original characteristic of living beings; i.e., not a first-base play.

It is important to emphasize that when we make a distinction between the observer's ascriptions and the phenomenon itself, it is not the point of the exercise to say that the observer's ascription is entirely groundless and unjustified. After all, the observer's biology is itself a natural system subject to natural laws and regularities. For example, in the famous Müller-Lyer illusion, frequently evoked as an argument for cognitive impenetrability of perception, we cannot but see the line with arrow-heads pointing inward as longer than the line with arrow-heads pointed outward. We see them this way even if we are told that they are in fact the same length. Of course, the perception is a mistake but it is not an arbitrary mistake. Rather, given the structure of our visual system, extensive experience with rectilinear environments (according to one theory) or the density of retinal pigmentation (according to another), the perception is entirely justified. However, it does not reveal anything real about the properties of the lines (their equal length).

Psychologists keep investigating perceptual illusions not because they are interested in inventing strange pictures. They accept the validity of the experience and ask questions of how the experience is produced which leads to uncovering brain mechanisms responsible for them, as well as interesting relations between the brain and the environment in which it is placed. They do not, however, use the experiences as *explanans* for some properties of the brain or the environment, i.e., they do not say "X is so because the lines are of different length" instead of "because the lines appear of different length". In a more quotidian example, the Sun can appear to move around the Earth but we do not explain the duration of the day by saying that it takes the Sun 24 h to orbit the Earth (not anymore anyway). Given our particular placement as terrestrial observers, the geocentric appearance is justified and has a real ground in the heliocentric organization of the solar system.

The idea to bear in mind is that once we manage to show the mechanisms that generate a determinate appearance in our observation, once we understand its ground, we cannot take said appearance as an operating element in the observed system, nor use it as an explanatory principle. Once explained, appearances can be recognized as valid phenomena of our observation, but not as properties of the observed systems, not even as alleged 'emergent properties'. In our previous example, the apparent geocentric organization of the solar system is not an 'emergent' property of the solar system; it is just an apparent property to the particular viewpoint of the terrestrial observer, which is something different.

In what follows we will try to show that both teleological and sensorimotor directedness, though erroneous, are somehow justified attributions in our observation of living beings. At the same time, and for the same reason, we will argue that, though justified, both forms of directedness are observer-relative ascriptions that, having no operational reality in living beings, cannot constitute the original ground for the game of content.

The alleged directedness of living beings' sensorimotor dynamics has been already questioned by Maturana on several occasions (Maturana 1980, 1981, 2003), and more recently by some enactivist critics (Villalobos and Ward 2015). The alleged teleological directedness, instead, we think, has not received yet enough attention (not at least within enactivist circles). For that reason in this paper we will dedicate more space to

develop, building mainly on Ashby's classical works on stability (Ashby 1960), a critique of teleological directedness, giving sensorimotor directedness only a general treatment.

Complex Stability Behind Teleological Directedness

Stability, in general, is a property that we can find in many systems, not only in living beings. In addition, all stable systems – as Ashby will show us soon – generate to greater or lesser degree a characteristic behavioral pattern that appears to be teleological (i.e., oriented at some goal or purpose). The more complex and richer the stability of the system, the stronger its teleological appearance. Living beings, as physicochemical systems that exhibit stability in a number of aspects (called homeostasis from Cannon (1932) onward), are not the exception to this rule but rather the most representative and strongest case, or so we will try to show in this section. The main idea is that the apparent teleological directedness of living beings is (1) precisely that, an appearance, and (2) that behind said appearance what exists is a complex form of stability.

To talk about stability we need a constant condition in the system, i.e., a steady state, and a perturbation that takes the system out of that condition; i.e., a disturbance. The constant condition or steady state provides the baseline with respect to which we can estimate whether or not the system exhibits stability. If the system, after being disturbed, spontaneously returns to its baseline, i.e., if it recovers its previous condition without external help or assistance, we say the system is stable (with respect to those variables or aspects under consideration). If the system does not return to its baseline, we say the system is unstable (with respect to those variables or aspects under consideration).

A system that remains in a constant condition or steady state is a system that, in most of the cases, is in some state of equilibrium (static or dynamic, mechanic or thermodynamic). That is, most of the steady states are equilibrium steady states. Non-equilibrium steady states are less common, but highly relevant for our understanding of living beings. As we saw in the previous section, and will see again soon, living beings, considered as thermodynamic systems, are dynamic patterns that exhibit stability in far-from-equilibrium conditions (i.e., they are dissipative structures).

The crucial point is that *every* stable system, when displaced from its baseline and then released, exhibits a line of behavior (direct or indirect, simple or complicated) that returns to its initial state (Ashby 1960). It is as if the system, in spite of disturbances, 'insisted' in maintaining its steady state and was 'directed' to it in its behavior once it was taken out of it. To take the simplest example of a static mechanic equilibrium, think of a cube lying on a horizontal surface. If tilted, it will return to the same 0° tilt of its initial resting position, thus showing stability. Similarly, a pendulum pulled to a certain height and released, will return to the center governed by the law of gravity. We can vary the direction of the disturbance, yet the pendulum will always recover its static equilibrium at the central position, thus showing stability. Ashby observes:

Noticing that the pendulum reacted with forces which though *varied* in direction *always* pointed towards the centre, the medieval scientist would have said "the pendulum seeks the centre". By this phrase he would have recognised that the

behavior of a stable system may be described as ‘goal seeking’ (Ashby 1960, p. 54, emphasis added).

All stable systems exhibit a typical behavioral pattern: no matter the way they are displaced from their steady state (the variability of the disturbances), they always return to it. It is this combination of variability (on the side of the behavior) with invariance (on the side of the steady state), that gives the idea of ‘goal-seeking’ in the system. The system, somehow, seems to have a fixed goal around which it is able to (‘adaptively’) vary its behavior according to the different circumstances. The medieval scientist would describe the pendulum behavior that conforms to such a pattern in teleological terms. We know, however, that pendulums are simple mechanical systems, absolutely blind and deaf to any purpose or goal. We know, and H&S would presumably agree, that in this case teleology is an appearance and that the behavior can be described by simple physical variables.

Why does the situation seem to change when we move to dealing with living systems? The answer, if we follow Ashby, has to do with the complexity of living beings as stable systems. The structural and functional properties that make living beings complex stable systems are many, but here we will address only a few: (1) the presence of feedback mechanisms, (2) the system’s dimensionality, (3) the organization of feedback mechanisms and (4) the system’s thermodynamic regime.

Feedback mechanisms, in simple terms, are functional loops in which a perturbation leads to a system’s response either in the same direction (in the case of positive feedback) or in the opposite direction (negative feedback). Cubes and pendulums reach their stable equilibrium states simply by ‘succumbing’ to physical laws, there is nothing like a functional circuit or servomechanism ‘helping’ them to compensate for the disturbances. On the contrary, as it is known from Cannon onward, living beings’ physiology is full of feedback mechanisms (especially negative ones) that automatically compensate transient disturbances. Yet, as it is also known, feedback mechanisms are present in many non-living systems too. Thermostats, automatic pilots, greenhouses, and the classical Watt governor are all examples of artificial systems that exhibit stability thanks to one or more negative feedback mechanisms. The presence of feedback mechanisms makes living beings different from cubes and pendulums, but not from thermostats, automatic pilots, Watt governors and similar devices.

The second feature we need to consider is the system’s dimensionality, i.e., the number of variables or aspects in which the system exhibits stability. Thermostats and Watt governors, for example, are stable systems that operate only in one dimension, namely temperature and speed engine respectively. A greenhouse operates typically in three or four – temperature, ventilation, humidity and luminosity. Living beings are stable systems of high multidimensionality; they operate in many dimensions at the same time (e.g., temperature, blood pressure, hydration, ph, hormone concentration, oxygen concentration, glucose concentration, electrolyte balance, etc.).

Things get a bit more complicated when stability variables are not merely numerous but also interconnected, i.e., when subsystems responsible for the maintenance of these variables cannot operate independently but instead affect each other. A greenhouse, for example, is able to keep temperature in check without its operation affecting the other three parameters. In living beings the stability of one of their subsystems, most of the times, compromises the balance of others (e.g., changes in the cardiovascular system

affect, sooner or later, the balance in the renal system, which in turn may affect the electrolyte balance of the body, etc.). Their high multidimensionality imposes a complex web of mutual restrictions among the variables, conditioning and complicating the global behavior of the system. Whereas the thermostat's behavior has to do uniquely with temperature stability and the greenhouse's behavior has a couple of more variables, the living being's behavior is the complex product of multiple interdependent stabilities running at the same time on multiple timescales.

Now, in principle, we might perfectly conceive of and build a stable artificial system of high multidimensionality equipped with many feedback mechanisms. Would that equate living beings' complexity as stable systems? Not yet. There is a third feature that must be taken into account, namely the order of the system's stability. Just like the organization of variables to be maintained, the organization of feedback mechanisms can also vary. When the circuit directly operates upon a determinate variable, we speak of a first-order feedback mechanism. When the circuit operates upon a determinate variable through the mediation of another feedback mechanism, we speak of a second-order feedback mechanism. A stable system that contains a second-order feedback mechanism exhibits a typical behavioral pattern that Ashby calls second-order stability, or ultrastability (Ashby 1960). An ultrastable system, thanks to the hierarchical organization of its feedback mechanisms, can exhibit highly complex forms of stability.

In living beings, the sensorimotor dynamics of an organism that is usually termed its behavior is an example of a first-order feedback mechanism. Sensory and motor surfaces of the organism constitute, through the mediation of the environment and the nervous system, a functional circuit that maintains a continuous correlation of activity. The potentially infinite variability of behaviors generated by sensorimotor circuits is limited by a second-order feedback mechanism that operates on essential physiological variables of the organism (e.g., oxygen concentration on blood, body temperature, hydration, etc.). In vertebrate animals, this second-order mechanism usually runs through some brain structure (e.g., basal ganglia or subcortical nuclei in general), whose activity connects with the sensorimotor mechanisms at different points.

Taking the example of pain reaction before an external stimulus, and simplifying a bit, every time the second-order circuit finds a level of sensory stimulation that goes beyond certain physiological threshold, it activates a step-mechanism that rearranges the activity of the sensorimotor circuit (the first-order feedback mechanism), thus generating a variation in the animal's behavior. This process continues until one of the behaviors so generated brings as a result the restoration of the sensory stimulation to its physiological values (e.g., avoiding the source of pain). In other words, the sensorimotor circuit is allowed to generate a considerable variability of behaviors, under the condition that none of those behaviors displaces the critical variable of the second-order circuits out of its physiological range.

Living beings' complexity as stable systems seems to lie in the fact that they comprise many ultrastable systems (at least one per every essential variable) running at the same time and in an interdependent way. They constitute what Ashby called multi-stable systems. Yet it was Ashby himself who showed that the basic principles of ultrastability could be equally observed in artificial systems. His famous Homeostat – a rather simple electromagnetic device equipped with second-order feedback mechanisms – is a clear demonstration of that (Ashby 1960). Again, one might conceive of and build a larger system, a Super Homeostat, containing many interconnected

ultrastable systems. Would we equate living beings' complexity with that kind of multi-stable system?

To understand the living beings' peculiarity as stable systems we need to consider a last and very important feature, namely their thermodynamic regime; i.e., the fact that, as we saw in the previous section, living beings are dissipative structures. What separates living beings from systems such as Homeostats, or even from more sophisticated versions of multidimensional multi-stable systems, is that they are constituted as physicochemical stabilizations in far-from-equilibrium conditions. All stable systems reviewed so far (from cubes to Homeostats) exist in equilibrium or near-equilibrium conditions. That is, their existence does not depend on a continuous exchange of matter and energy with the environment. If you cut the energy supply to a thermostat, it will cease to operate, but it will not disintegrate (you can turn it on again and the system will be operative). Dissipative structures exist in non-equilibrium conditions, and if they go to thermodynamic equilibrium they not only cease to operate, they disintegrate (Bickhard 2007).

All dissipative structures exhibit a greater or lesser degree of stability around their particular thermodynamic, non-equilibrium, viability regions, returning to their patterns of material and energetic exchange with the environment when disturbed and thereby conserving their identity. When observing this process occurring in a living being, the typical image is that of a system 'struggling' to survive and maintain its integrity (e.g., the Spinozan 'conatus'). The teleological appearance is particularly strong in this case, since among dissipative systems, living beings are the only ones equipped with a set of interconnected second-order feedback mechanisms.

As far as we know, living beings seem to be the unique multi-stable dissipative systems in Nature. For example, if in an animal the intake of energy and matter is diminished to some degree (e.g., it has not eaten or drunk anything for a relatively long period), and the system, because of this, is displaced from its steady state as a thermodynamic system (i.e., essential variables such as hydration or glucose concentration deviate from their optimal values), at some moment the system will automatically operate, through one or more ultrastable mechanisms, the reactivation of its sensorimotor circuit. From that moment, the ultrastable dynamics held between the second-order feedback mechanism (which operates upon essential variables) and the first-order feedback mechanism (which operates upon the correlations of sensorimotor activity), will iterate until the behavior of the system brings as a result an intake of matter and energy such that the essential variables return to their optimal values. In this complex process, what the observer will see from the outside is that the animal gets up, walks to the kitchen and drinks some water, all this with the 'purpose' of restoring its physiological levels of hydration.

Given the high complexity of the living being's stability, one could say that the observer is more than justified to attribute teleological directedness to its behaviour. As in the case of the solar system and its geocentric appearance, nonetheless, it remains the fact that, though justified, the alleged teleology is just an appearance in the eye of the beholder and not an intrinsic property of the living being.

Functional Closure Behind Sensorimotor Directedness

Suppose, H&S are convinced that our characterization of living beings is correct so far; that the teleological directedness of living beings is just an appearance, and that the real

underlying mechanism is stability. They could still maintain that the possession of the sensorimotor system *by the organism* and the fact that this system is employed in the maintenance *of the organism*, justifies the claim that the organism is aimed at certain features of the environment. Actually, it would be hard to deny that when one sees a living being in its environment, what one sees is an organism receiving certain stimulation *from* the environment and responding in turn through some action executed *upon* the environment. Living beings' dynamics looks clearly directed at the external world. Can we be mistaken about that? According to Maturana's autopoietic theory, we not only can but actually are mistaken about that (Maturana 1980, 1975, 2003). As in the case of the geocentric appearance, the sensorimotor directedness of living beings would be a 'normal' and expectable appearance that has to do, on the one hand, with the intrinsic organization of the observed system, and on the other, with our particular position as observers in relationship to the said system.

In the previous section we talked about the sensorimotor system as a first-order feedback mechanism, and we said that feedback mechanisms are basically functional loops. Now we have to look closer the organization of the sensorimotor system as a functional loop. Feedback mechanisms are functional loops because they are organized as closed circuits of activity, as circular sequences of causes and effects. In the case of the sensorimotor system, the sensory surfaces connect with the motor surfaces through the mediation of the nervous system, and the motor surfaces connect with the sensory surfaces through the mediation of the environment. For instance, what the animal encounters as stimulation at the level of its retina at a given moment depends on the particular position and movement of its body at that moment, position and movement that are determined by its motor activity. At the same time, the motor activity of the animal at a given moment depends, in part, on what happens at the level of its retinal stimulation.

In a circular network of sensor-effector correlations like this, strictly speaking, there are not inputs or outputs, because the system, *from the functional point of view*, is always closed on itself.

The sensory and the effector surfaces that an observer can describe in an actual organism, do not make the [...] system an open [...] network because the environment (in which the observer stands) acts only as an intervening element through which the effector and sensory [surfaces] interact completing the closure of the system (Maturana 1975, p. 318).

The key point is that the environment, in which we observers stand, plays a role that for the sensorimotor system itself proves to be functionally transparent, indistinguishable from any other functional step of its circular dynamics. The sensorimotor system, considered not in its anatomical dimension but from the functional point of view, has the environment as one more of its functional steps, i.e., as something internal, not as an external element before it to which the sensorimotor dynamics can be directed. The directedness of the sensorimotor system is an appearance resulting from our position as observers.

An observer that sees an effector/sensor correlation as an adequate behavior does so because he or she beholds the organism in the domain of structural coupling in

which the distinguished behavior takes place in the flow of its conservation of adaptation. The organism in its operation [, nonetheless,] does not act upon an environment; the environment exists only for an observer (Maturana 2003, pp. 102–103).

As in the case of stability, there is a real phenomenon to be described here, which includes the functional closure of the sensorimotor system and the observer positioned in a particular place of the loop. We treat the effector activity as directed to the external world because we are standing outside the organism, in the gap between its sensors and effectors. Were we placed in a different point of the functional circuit, e.g., in a synaptic cleft between two neurons, we would treat that functional gap as the ‘world’ towards which the rest of the circuit is directed (Maturana 1975). In other words, the alleged sensorimotor directedness of the organism proves to be an observer relative notion, not an intrinsic feature of the observed system.

The functional closure of the sensorimotor system, if we follow Maturana’s analysis, renders any directedness (content-involving or contentless) an arbitrary convention introduced by the observer; i.e., a second-base ascription that, again, cannot work as the original ground for the game of content.

Implications for the Theory of Content

If the reader has reached this point, she might be asking herself: where does all this leave us with respect to the origin of content? We have devoted our section 4 to making the point that the features characteristic of living systems as multi-stable systems help explain not merely their peculiarity but also why (Ur-)intentionality is *attributed* to them in such an intuitive manner. Accordingly, we would like our takeaway message to be: Ur-intentionality is a second-base notion, not a first-base ground that can help us understand the origin of content.

To recall, H&S delegate the second base to Neo-behaviorists who claim that content is a useful description of the system that helps us understand its behavior but does not have reality in the system itself. H&S observe that left to their own resources, second basemen are unable to explain how the ability to adopt an intentional stance (which is required for making content ascriptions) emerged on the natural scene. Without providing such an explanation, Neo-behaviorism ends up presupposing content instead of illuminating its origin.

Ur-intentionality is offered as an alternative to this merely-ascribed kind of contentful intentionality. It is intentionality of a leaner kind that is able to explain adaptive behavior of contentless organisms and one that, given the right conditions, can bloom into the fully fledged intentionality of the second base type. What we have been trying to show, however, is that Ur-intentionality itself is a feature ascribed to behaving organisms and not their intrinsic property. If this is so, then analogously to the Neo-behaviorist predicament, it ends up presupposing certain ascriptional abilities instead of explaining their emergence and something else is required to ground both Ur-intentionality and contentful intentionality of the old days.

Suppose we accept this conclusion. Does it mean we should drop our bats and gloves and leave the field altogether? Or does it mean perhaps that there is no content or

intentionality all the way up? For reasons of space, we cannot fully answer this challenge in the current article. However, in the remaining few paragraphs we will provide a sketch of what our reply would contain.

We believe that we would need to start by asking: what are the reasons for postulating the existence of Ur-intentionality in the first place? We interpret H&S as thinking there are two: (1) a story on the origin of content, cannot have it emerge out of thin air, so to say; (2) if content arises together with language and culture, we need to provide an explanation of the cognitive abilities of contentless creatures – abilities that enable them to act in the world, interact with each other and eventually evolve cultural transmission and language.

We think the first reason is unnecessary and often leads to problems with drawing arbitrary boundaries and making anthropomorphic projections. When we say that bacteria searching for sucrose exhibit Ur-intentionality we create an impression that we understood something – that here, in this behavior the human property that we really care about is born. All we need to do now is tell a story on the evolution of some further features of the human mind that transform this seed into a full specimen. We have connected ourselves with bacteria but why not push further? Why not attribute Ur-Ur-intentionality to autocatalytic cycles in a petri dish or stones on a sidewalk? The price being, however, that in the search for the origins of our features, we might overlook the features that make rocks and bacteria special in their own rights, features that might let us see how we are *like them*, not how they are *like us* and as a result let us understand ourselves better.

Objecting to the second reason properly requires a separate article but here we wish to point out that the emergence of linguistic behaviors can be cashed out in terms consistent with Strict Naturalism. In fact, the beginnings of such an account have been provided by Maturana (1978). It requires us, however, to conceive of living beings as we have presented them above – as very special and unique but essentially machines composed of chains of deterministic processes. Everything that the system does is determined by its structure at a given moment, not goals or desires about states in the future and not selecting between features of the environment that harm or benefit the organism.

Maturana (2003, p. 72) provides an illuminating metaphor for the life process:

It is said that a boat is drifting when it slides floating on the sea without rudder and oars, following a course that is generated moment after moment in its encounter with the waves and wind that impinge upon it, and which lasts as long as it remains floating (conserves adaptation) and keeps the shape of a boat (conserves organization). As such a drifting boat follows a course without alternatives that is deterministically generated moment after moment in its encounters with the waves and the wind.

He insists that the organism is similarly in a drift during its ontogeny following a course “in the braiding of its internally generated structural dynamics with the structural changes triggered in it by its recurrent interactions with the medium”. Any selection or directedness is only in the eye of the observer who can envision how things could have been different or what their ultimate outcome turned out to be. And the same goes for phylogenies – histories of structural changes in the species.

The behavior is adaptive with respect to the environment (the internal structures extensionally covary with external states of affairs, in the H&S parlance) because through recurrent interactions the organism has undergone a process of structural coupling to its niche. The system's structure is plastic, i.e., in an organism with a nervous system the patterns of neural responses can change. If they do, the system enters new interactions with a new structure. In this flow of constant change in interaction with the environment what emerges is "the spatiotemporal coincidence between the changes of state" (Maturana 1975, p. 321) of the elements participating in the coupling.

Now, the key point is that one of these participants can be another agent, not just 'passive' environment. In recurrent structural interactions two agents can become coupled to each other, serving as each other's perturbations and thereby affecting each other's behavioral trajectories. This, on autopoietic account, constitutes the basis for language – a domain in which both content and intentionality have their place (where we are in agreement with H&S's third base).

Of course, a lot more remains to be said about how precisely this is accomplished but the same goes for any Ur-intentionality-based account. We believe, however, that autopoietic framework, which is consistent with Strictly Naturalistic science that deals with structure-determined systems, is a viable alternative to be pursued.

References

- Asby, W. R. (1960). *Design for a brain*. London: Chapman & Hall.
- Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Cambridge: MIT Press.
- Bickhard, M. H. (2000). Autonomy, function, and representation. *Communication and Cognition Artificial Intelligence*, 17, 111–131.
- Bickhard, M. H. (2007). Mechanism is not enough. *Pragmatics and Cognition*, 17(3), 573–585.
- Cannon, W. B. (1932). *The wisdom of the body*. New York: The Norton Library.
- Cummins, R., Blackmon, J., Byrd, D., Lee, A., & Roth, M. (2006). Representation and unexploited content. In G. MacDonald & D. Papineau (Eds.), *Teleosemantics* (pp. 195–207). Oxford: Oxford University Press.
- Hutto, D. D. (2008). *Folk psychological narratives*. Cambridge: MIT Press.
- Hutto, D. D. (2011). Philosophy of mind's new lease on life: autopoietic enactivism meets teleosemiotics. *Journal of Consciousness Studies*, 18(5–6), 44–64.
- Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge: MIT Press.
- Hutto, D. D., & Satne, G. (2014). The natural origins of content. *Philosophia*.
- Ji, S. (2012). *Molecular theory of the living cell: Concepts, molecular mechanisms, and biomedical applications*. New York: Springer.
- Juarrero, A. (2002). *Dynamics in action: Intentional behavior as a complex system*. Cambridge: MIT Press.
- Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. London: MIT Press.
- Maturana, H. (1975). The organization of the living: a theory of the living organization. *International Journal of Man-Machine Studies*, 7, 313–332.
- Maturana, H. (1978). Biology of language: The epistemology of reality. In G. Miller & E. Lenneberg (Eds.), *Psychology and biology of language and thought: Essays in honor of Eric Lenneberg* (pp. 27–64). New York: Academic.
- Maturana, H. (1980). Biology of cognition. In H. Maturana & F. Varela (Eds.), *Autopoiesis and cognition: The realization of the living* (pp. 5–58). Dordrecht: Reidel Publishing Co.
- Maturana, H. (1981). Autopoiesis. In M. Zeleny (Ed.), *Autopoiesis: A theory of living organization* (pp. 21–33). New York: North Holland.

- Maturana, H. (2003). The biological foundations of self-consciousness and the physical domain of existence. In N. Luhmann, H. Maturana, M. Namiki, V. Redder, & F. Varela (Eds.), *Beobachter: Convergenz der erkenntnistheorien?* (pp. 47–117). München: Wilhelm Fink Verlag.
- Millikan, R. G. (2004). *Varieties of meaning: The 2002 Jean Nicod Lectures*. Cambridge: MIT Press.
- Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems: From dissipative structure to order through fluctuations*. New York: Wiley.
- Prigogine, I. (1980). *From being to becoming: Time and complexity in the physical sciences*. USA: Freeman and Company.
- Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature*. Canada: Bantam Books.
- Saborido, C., Mossio, M., & Moreno, A. (2011). Biological organization and crossgeneration function. *The British Journal for the Philosophy of Science*, 62, 583–606.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge: Harvard University Press.
- Ulanowicz, R. E. (2009). *A third window: Natural life beyond newton and darwin*. PA: Templeton Foundation Press.
- Ulanowicz, R. E., & Hannon, B. M. (1987). Life and the production of entropy. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 232(1267), 181–192.
- Villalobos, M. & Ward, D. (2015). Living systems: autopoiesis, autonomy and enaction. *Philosophy and Technology*, 28(2), 225–239.