AVANT, Vol. X, No. 3/2019
ISSN: 2082-6710 avant.edu.pl/en
DOI: 10.26913/avant.2019.03.11

Unsupervised Learning and the Natural Origins of Content

Tomasz Korbak
Institute of Philosophy and Sociology, Polish Academy of Sciences
University of Warsaw
tomasz.korbak@gmail.com

Received 26 February 2019; accepted 21 August 2019; published 12 September 2019.

Abstract

In this paper, I evaluate the prospects and limitations of radical enactivism as recently developed by Hutto and Myin (henceforth, "H&M") (2013, 2017). According to radical enactivism, cognition does not essentially involve content and admits explanations on a semantic level only as far as cognition is scaffolded with social and linguistic practices. I investigate their claims, focusing on H&M's criticism of the predictive processing account of cognition (dubbed the *bootstrap hell* argument) and their own account of the emergence of content (the *natural origins of content*). I argue that H&M fail on two fronts: unsupervised learning can arrive at contentful representations and H&M's account of the emergence of content assumes an equivalent bootstrapping. My case is illustrated with Skyrms' evolutionary game-theoretic account of the emergence of content and recent deep learning research on neural language models. These arguments cast a shadow of doubt on whether radical enactivism is philosophically interesting or empirically plausible.

Keywords: hard problem of content; radical enactivism; predictive processing; neural language models; deep learning; bootstrap hell; semantic information.

1. Introduction

In this paper, I evaluate the prospects and limitations of radical enactivism as recently developed by Hutto and Myin (henceforth, "H&M") (2013, 2017). According to radical enactivism, cognition does not essentially involve content and admits explanations on a semantic level only as far as it is scaffolded with social and linguistic practices. Basic minds, i.e. phylogenetically and ontogenetically early cognition, are to be explained in

terms of dynamics, sensorimotor couplings with environment and contentless goal-directedness. This is because there are, H&M claim, serious philosophical problems with applying a semantic-level vocabulary of representations, models, and computations: it is troublesome to give a non-circular account of how content emerges in the natural world. This worry is known as the hard problem of content (HPC).

Numerous authors have argued this view to be indefensible, first because the HPC argument is flawed, and secondly, because H&M fail to provide a positive research program for cognitive science (e.g. Alksnis, 2015; Harvey, 2015; Korbak, 2015). My focus in this paper will be to review these two worries and evaluate whether H&M's most recent account (in their 2017 book *Evolving Enactivism*) does address these worries and can be defended against them.

The HPC is most relevant for modern neuroscience as a voice against predictive processing approaches to cognition that have recently been gaining momentum (Hohwy, 2013; Clark, 2016). I will therefore focus my discussion on the "bootstrap hell" argument, a special case of HPC, putatively demonstrating the failure of predictive processing. After that, I will reconsider H&M's own story of the emergence of content in mature, socioculturally embedded minds: the "natural origins of content," and how well it fares. My conclusion is that the natural origins of content story itself must get out of its bootstrap hell, and if it does, the HPC is easily solvable, and the natural origins of content will lie much earlier in natural history than H&M claim. Therefore, radical enactivism, the claim that basic minds are contentless, is either false or trivial.

2. Facing Backwards on the Hard Problem of Content

Hutto and Myin argue that minds are not essentially contentful, because content-involving philosophical accounts of mind fall prey to the HPC. The HPC argument can be reconstructed as follows (Korbak, 2015, p. 90):

> (T1) Ontological commitments in cognitive science must respect explanatory naturalism.
> (T2) Linguistic activity is out of the scope of basic minds.
> (T3) Having content implies having certain satisfaction conditions, which determine intension and extension (if it exists).
> (T4) Every possible theory of content that attributes content to basic minds fails to respect either (T1), or (T2), or (T3).
> (T5) Having content is constitutive for being a representation.

(T1)–(T5) jointly imply that representationalism, i.e., the view that cognition essentially involves representation, is false, which follows directly from (T4) and (T5) being true. It would be of major importance for cognitive science, if that were the case.

To carry out their argument, H&M focus on the most controversial premise, (T4), and review several existing theories of content—including Dretske's indicator semantics

(Dretske, 1983) and Millikan's teleosemantics (Millikan, 1984). Naturalistic accounts of content aim at reducing content to something ontologically simpler, for instance, natural laws, Shannon information, or biological proper function. H&M claim that the base needed for a reductive explanation of content must include language use (or, in general, a sociocultural scaffolding of shared conventions) that appears late in phylogeny and ontogeny. H&M's dismissal of existing accounts is fairly premature (Miłkowski, 2015), but I will not explicitly pursue this line of criticism here. Instead, I will argue against (T2). First, I will do this by arguing that H&M's definition of "linguistic activity" is so broad that it covers any and all transmission of semantic information. Second, I will argue that the transmission of semantic information can be found very early in natural history.

An important point in H&M's criticism is that covariance does not constitute content: growth rings of a tree may be systematically correlated with its age, but that doesn't mean they are *about* age. Starting from this remark, they accuse most of contemporary philosophy of an equivocation fallacy: not distinguishing between Shannon information[1] and semantic information. This is indeed an important distinction.[2] Surely, the distinction counts for living systems: maintaining the flow of information in the body is costly and can reasonably be expected to serve some aim rather than being a by-product. If Shannon information "did not carry information about anything, nobody would be in the business of communicating it." (Miłkowski, in press). For the purpose of this paper, I assume that what distinguishes semantic information from Shannon information is that semantic information gives rise to satisfaction conditions (it presents the world as being such-and-such), which the world can meet or fail to meet. The possibility of the latter—*mis*information—is an important desideratum for a theory of content.

3. The Bootstrap Hell Argument

Bootstrap hell is a special case of the HPC that is supposedly faced by predictive processing accounts of cognition. Predictive processing is a family of approaches in neuroscience, psychology, and philosophy that seek to explain cognition in terms of hierarchical generative models and minimization of prediction error. Wanja Wiese and Thomas Metzinger (2017) list seven claims that are usually shared by various flavors of predictive processing: (1) recognizing the role of top-down information processing in cognitive systems,

[1] H&M don't appeal to the formal concept of Shannon information, but I believe exposition of their claims benefits in clarity from employing this concept: The mathematical theory of communication (Shannon, 1948) explicates H&M's concept of information-as-covariance and also considers the entropy of information source and channel noise, which are absent in H&M's treatment.

[2] Several authors have, however, argued that the distinction between Shannon information and semantic information is either oversimplified (Rathkopf, 2017), or misleading, because Shannon information can do most of the work usually attributed to semantic information (Bergstrom & Rosvall, 2011; Martinez, 2019). For the purpose of this paper, I accept the received view that the distinction is valid.

(2) maintaining that cognition involves modeling distributions of random variables deployed by (3) hierarchically organized generative models, and used for (4) prediction. These generative models are (5) fine-tuned to reduce prediction error (6) in a Bayes-optimal fashion. Finally, predictive processing also claims (7) motor control is explainable in terms of Bayesian inference.

Predictive processing offers an ambitious integrating account of cognition that has received considerable interest in many areas of cognitive science, including computational psychiatry (Adams, Huys, & Roiser, 2015), affective neuroscience (Barrett, 2018), consciousness studies (Seth, Suzuki, & Critchley, 2012) and developmental robotics (Tani, 2017). Philosophically, predictive processing puts forth a radically new image of perception, learning, imagination, and action intricately coupled and deeply embodied in biological autonomy while self-organizing around prediction errors (Clark, 2016). Radical enactivists, however, still accuse predictive processing of not being radical enough.

It is the commitment to predictions produced by generative models that bothers radical enactivists, because the notion of prediction essentially involves content: The future being predicted is predicted as being such-and-such, and predictions sometimes fail to come true. In predictive processing, content is mostly produced top-down rather than received from sensors and consumed by the system; prediction errors are meaningful only relative to the original predictions. According to Andy Clark "the prediction task is […] a kind of bootstrap heaven" (Clark, 2016, p. 19) as the content of predictions is iteratively refined based on errors, without the need for direct access to the hidden causes of the sensory signals. Hutto worries that this heaven may actually turn out to be hell, since the brain (or the body) lacks resources to give rise to contentful predictions in the first place. "If minds are in principle forever secluded from the world how do they come by contents that refer to, or are about, inaccessible hidden causes and external topics that they putatively represent in the first place?" (Hutto, 2018, p. 13). While the content of subsequent predictions may be argued to be a product of Bayes-optimal integration of prediction errors with previous predictions, there remains the problem of the first prediction in the chain.

This forms the objection at the heart of the bootstrap hell argument: What determined the content of the first prediction that a cognitive system produced? As Hutto vividly states:

> It is one thing to create a large fire from a smaller one, and in certain conditions that can be quite a difficult business. It is quite another thing to create a fire from scratch with only limited tools, of, say, flint and steel, especially when conditions are not favourable. (Hutto, 2018, p. 10)

The hidden premise in this objection is that it is problematic for content to emerge spontaneously in an adaptive system. This is why the bootstrap hell is a special case of HPC. One reason, however, why I find the bootstrapping formulation interesting is that H&M's own account of the natural origins of content in human language seems to involve a bootstrapping process (see section 6).

H&M focus their efforts on showing that existing accounts fail to account for the natural origins of content. Rather than defending Dretske's or Millikan's, I prefer to mount an argument against radical enactivism by assuming a minimalistic account of content developed by Brian Skyrms (2010).[3] This account explains the emergence of content in terms of a game of message passing between a sender and a receiver (Lewis, 1969). The sender sees certain inputs and has certain preferences about the receiver's behavior (generally relative to the input) and the receiver is free to act based on the message (but does not see the input). According to Skyrms, the content of a message consists in how it affects the probability distribution over actions the receiver may undertake (Skyrms, 2010, p. 41).[4] The strategies of both the sender (which message to send) and the receiver (a stochastic mapping from the message to an action) are subject to evolution and/or learning, which is shaped by reward for the joint performance of the sender–receiver system. This is why "[i]nformational content evolves as strategies evolve" (Skyrms, 2010, p. 35) and the gap between natural meaning (in the sense of Grice (1957) or what H&M dub "lawful covariance") and conventional meaning dissipates: conventionality enters the picture as there are degrees of freedom in the sender-receiver system under its fitness landscape, i.e., multiple possible codes for controlling the receiver. There is no need for a social scaffolding other than the sender–receiver system itself to get communication off the ground.

H&M will be quick to repeat their slogan that covariance does not constitute content. But while informational content *sensu* Skyrms is founded on Shannon information, it is something more. This is because *how* the probabilities (over receiver's action) change is something more than *how much* they do, i.e., the quantity of Shannon information. The former has well-defined satisfaction conditions, can fail to affect the receiver, and is subject to error correction and optimization (e.g., learning in a short timescale or development and evolution in a long one). Skyrms offers a proof-of-concept of an account of content solving the HPC.[5]

The power of Skyrms' account lies in the fact that it solves more than armchair philosophers' problems. Both formal, (evolutionary) game-theoretic analyses (Shea, Godfrey-Smith, & Cao, 2017) as well as computational simulations (Bouchacourt & Baroni, 2018) show that a wide variety of natural phenomena can be modeled in terms of sender–receiver dynamics. Such phenomena include intra-cellular signaling, animal communication, representation learning in neural networks and brains, and evolution of human language. In the next section, we will quickly illustrate how Skyrms' account can shed light on the training of state-of-the-art deep neural networks.

[3] Arguments based on Skyrms' will work with other accounts of semantic information sharing the emphasis on action guidance, receiver-relativity and optimization. Numerous accounts starting from these principles have emerged in recent years, including those of Bickhard (2009), Dennett (2017), and Kolchinsky and Wolpert (2018).

[4] For completeness, Skyrms also distinguishes a more traditionally flavored descriptive content vector: The descriptive content of a message is the change in probability distribution over the inputs of the sender given the message the sender has sent. The following discussions focused on the first vector, which can be associated with imperative content.

[5] For a recent defense of Skyrms' account of content, see (Isaac, 2019).

4. Unsupervised Learning in Artificial Neural Networks, Brains, and Societies

A *language model* is a generative model that assigns probabilities to sequences of words (Jelinek & Mercer, 1980). By the chain rule of probability calculus, this objective can be reformulated as predicting the next word in a sentence conditioned on a few previous words. Language models are widely used in natural language processing, powering technologies such as information retrieval, speech recognition or spell-checking. Language models are instances of unsupervised machine learning: They only require the words in the training set to be linearly ordered (i.e., coming from a contiguous text), not assuming any labels to be assigned.[6] Language models are usually trained to exploit statistical patterns found in text available on the Internet. It turns out, however, that a great deal of lexical information, as well as morphology, syntax, semantics, and pragmatic conventions can be learned from unlabeled data, thus laying a path towards artificial general linguistic intelligence (Yogatama et al., 2019).

More importantly, language modeling is an active area of research in deep learning, because representations found to be useful for language modeling are surprisingly reusable for other tasks, including syntactic parsing, question answering, and machine translation (Peters et al., 2018; Radford et al., 2019). Therefore, representations learned by a *neural language model* (a deep neural network trained on language modeling) can be argued to have semantic content encoded by a few first layers and exploitable by subsequent layers. This can be seen quite concretely when considering a phase space spanned by weights of a particular layer of a deep neural network. The activation of a (layer of) a neural network processing each word corresponds to a particular point in this space, a vector known as a *word embedding*. Word embeddings curiously encode semantic and syntactic relations between words in terms of geometric relations in the phase space, e.g., synonyms will be close to each other (in terms of cosine similarity, or the normalized angle between a pair of points). Simple cases of lexical inference can also be replicated by word embedding arithmetic, for instance the word embedding for "Paris" minus "France" plus "Poland" happens to be very close to "Warsaw." This is usually interpreted as evidence for rudimentary commonsense knowledge encoded in word embeddings (Mikolov et al., 2013). It's no surprise word embeddings make useful features for a wide array of machine learning tasks; as of 2019, they form part of the standard toolbox of a natural language processing engineer. Yet according to H&M, since there is no sociocultural scaffolding allowed, deep neural networks are doomed to fall prey to the HPC. They are either (counter-intuitively) contentless or metaphysically impossible. I close this section by arguing that deep neural networks actually acquire genuinely contentful representations in a process of unsupervised learning, analogous to the unsupervised learning posited by predictive processing. If that is so, there is nothing preventing the brain itself to acquire contentful representations via unsupervised learning, which renders the HPC argument ill-posed.

[6] Unsupervised learning is learning with a loss function that depends only on the inputs to the model (and statistical patterns they exhibit) and does not depend on any extrinsic information. The distinction between supervised and unsupervised learning is somewhat arbitrary and there are many shades of gray in between. Both language modeling and time series prediction are, however, usually considered to be unsupervised learning.

Let me point out that, mathematically, the problem of language modeling is pretty much equivalent to the problem faced by the brain according to predictive processing. The goal of a language model is to predict the next word. The goal of the brain is to predict the next sensory input. In both cases, the loss function to be minimized is expected surprise, i.e., average negative log probability of training data given the model.[7]

So how does a neural language model learn meaning *ex nihilo*? What determines the word embedding guiding the first prediction (for the first word in the training set)? Pure noise. It is standard practice to initialize layer weights randomly (sampling from a zero-centered Gaussian with relatively low variance). The network will then most likely predict an approximately uniform probability distribution over the whole vocabulary. Now let us interpret the aforementioned layer as the sender and the following one a receiver. Note that even at this point the sender meets Skyrms' criteria for informational content. It's just a very uninteresting content, of little use for the receiver. This will, of course, change in the course of training: The sender's messages will gradually drive the receiver to contribute to moving the probability mass over the vocabulary to the right place.

Recall that an important (and uncontroversial) assumption behind HPC is that (T3) having content implies having certain satisfaction conditions, which determine intension. Intension specifies what it takes to be referred to by a content vehicle. This condition is known as *aboutness* or intensionality-with-an-s.[8] My point is that representations of a trained neural language model will exhibit intensionality-with-an-s: Word embeddings have satisfaction conditions by virtue of the influence they have on the predictions of the neural network (the meaning of a given word embedding is how it changes the probability distribution over the next word to appear in text). This kind of meaning may not be fully aligned with the conventional meaning we ascribe to respective words as it is natural meaning stemming from word cooccurrence patterns in the training set exploited by the neural network. However, it provides each information vehicle (a set of weights) with an intension (the change in probability distribution over the next word to appear in text).

One possible objection that a radical enactivist could raise at this point is that word embeddings are meaningful only by virtue of the intensionality of humans. H&M make a heavy use of this strategy when accusing existing naturalist accounts of content of circularity (e.g., Hutto & Myin, 2013, p. 70). In our setting, the objection could be formulated as follows: word embeddings are not intrinsically contentful, but only appear to be contentful to

[7] There are important differences in architecture and optimization procedure between (deep) neural networks and hierarchical Bayesian models in computational neuroscience. Drawing the analogy between the brain and language models also requires a few words of caution: Language models are not properly embodied and situated and are not supposed to model important aspects of the brain. They can also be argued to fall short of the task of human language use, because they lack a feedback loop with extra-linguistic reality. Language modeling still seems to be on the first step of Judea Pearl's ladder of causation (Pearl & Mackenzie, 2018), because language models have no means of engaging in dialogue with a text. These are valid concerns for artificial intelligence researchers but are not immediately relevant for the argument mounted here.

[8] This is to avoid confusing intensionality with intentionality.

human engineers who already know the lexical meaning of corresponding words and project their linguistic competence to contentless neural networks. If that is so, ascribing content to word embeddings presupposes human brains to be content-involving rather than helping to explain how brains come to be content-involving. Therefore, the whole argument mounted in this paper would be circular: the emergence of content in neural language models would still require a social scaffolding constituted by machine learning engineers.

While word embeddings are receiver-relative in the sense of being sensitive to a particular neural language model architecture, training set, and training procedure, they do not presuppose an intensional receiver other than a mindless optimization process of a loss function. On the contrary, word embeddings are meaningful for neural language models but quite opaque for humans; special tools need to be developed for interpreting them (Pelevina, Arefiev, Biemann, & Panchenko, 2016). The meaning of word embeddings instantiates natural meaning (by being based on word co-occurrence patterns in the training set) harnessed specifically for a given loss function (predicting the next word). Moreover, there is nothing inherently social in the loss function other than requiring several layers of a neural network to cooperate.

As H&M like to begin each chapter with lyrics from The Beatles, I should probably have started this section with "No hell below us, above us only sky." There is simply no deep problem with bootstrapping unsupervised learning. An unsupervised learner will obviously fail miserably at first, but all it takes to support the learning process is having degrees of freedom in the system and some feedback about its performance available for the system. That's how learning in humans (and other machines) works.

5. Natural Origins of Content

H&M "are not content-deniers; they do not embrace global eliminativism about content" (Hutto & Myin, 2017, p. 121). They claim that while basic minds don't usually deal with contentful representations, phylogenetically and ontogenetically mature cognition—when immersed in language and other symbolic forms of culture—does. Radical enactivism thus faces a problem similar to the HPC: how to account for the emergence of content. The problem is supposedly manageable only because sociocultural scaffoldings, arising late in evolution, are the right resources to solve the problem.

Radicalizing Enactivism (Hutto & Myin, 2013) was widely criticized for lacking a positive story of what precisely gives rise to content in mature minds (e.g. Alksnis, 2015; Harvey, 2015; Korbak, 2015). *Evolving Enactivism* was supposed to fill this gap and tell the whole story of how *basic minds meet content* (according to the subtitle). It is thus slightly dissatisfying that the relevant Chapter 6 focuses mostly on finding gaps in arguments mounted by radical enactivism's adversaries. H&M find most of these arguments accusing radical enactivism of violating the principle of evolutionary continuity by assuming a difference of kind, not degree, between basic (contentless) and mature (content-involving) minds.

But is discontinuity the only problem faced by the natural origins of content program? Let me use this section to comment on an argument I put forth against the story in *Radicalizing Enactivism* (Korbak, 2015). According to the scaling down objection (as H&M dub it), whatever it is that makes healthy human adults capable of exchanging content should also make the ability appear in simpler adaptive systems, including the endocrine system, intracellular signaling pathways and gene expression regulation mechanisms in single-cell organisms. This is because the repertoire of kosher *explanantia* is so limited: H&M claim language use to span the scaffolding for content to thrive on, but they are forced to assume a very minimal concept of language—language being a tool for shared action guidance. Only then can they avoid the circularity of assuming a traditional account of language as essentially contentful. There is nothing wrong with such an account of language (it is treated quite seriously in experimental semiotics, e.g., Pattee & Rączaszek-Leonardi, 2012) but from this view, every bacterium also counts as a language user, hence a mature mind. The class of basic (contentless) minds is then (by our understanding of life on Earth) empty, because exchanging content is necessary for biological regulation. In other words, in H&M's own characterization of language as a tool to joint action, language is to be found very early in phylogeny, thus building the scaffolding for the very early emergence of content as well.

The scaling down objection is not about evolutionary continuity *per se*, as H&M cite it to be (Hutto & Myin, 2017, pp. 131–132). The qualitative change entailed by the natural origins of content story is not the worry here. The worry is that there can't be a qualitative change, because H&M's sociocultural scaffolding starts to apply much earlier down the tree of life. To avoid this objection, H&M must establish criteria for content emergence that (i) do not circularly assume the flow of informational content, and (ii) are indeed found only late in phylogeny and ontogeny. Let us see what progress H&M have made on this front:

> For the sociocultural emergence of content we need to assume that our ancestors were capable of social processes of learning from other members of the species and that they established cultural practices and institutions that stabilized over time. […] The capacities in question can be understood in biological terms as mechanism through which basic minds are *set up to be set up* by other minds and *to be set off* by certain things. (Hutto & Myin, 2017, p. 139, emphasis original)

It requires a grain of good will to interpret the social as not invoking persons with contentful attitudes, thus satisfying desideratum (i). But if the social means *multi-agent*, what prevents societies of cells (also known as organisms) or societies of proteins (also called cells) from agreeing upon the content of messages they pass to each other? We know they do: Orchestrating the workings of a complex system requires high-throughput flow of information and content being used for this orchestration evolves like any other trait. Sometimes, when the sender misinforms the receiver, the system suffers a pathological state, as in the case of cancer, autoimmunity or diabetes. Luckily, understanding the mechanisms involved in the flow of semantic information through a complex living system, we are sometimes able to intervene and transmit our own message. This is how depression, Parkinson's and endocrine diseases are treated, and how genetically modified organisms are obtained. Sometimes

the sender and receiver have partially misaligned goals and the sender may send a dishonest message. Such cases are frequent in animal sexual behavior (Skyrms, 2010, ch. 6).

I dare to go further and argue that H&M's natural origins of content is more or less equivalent to Skyrms' account of the evolutionary dynamics of informational content. It is the receiver that is *set up to be set up* by the sender, and it is the sender's role to *be set off* by relevant events in their niche. One crucial difference is that Skyrms' account is more mature and formally elegant, for instance, the concept of "established cultural practices and institutions that stabilized over time" can be made less hand-wavy in terms of an evolutionarily stable strategy.

One could argue that I'm not being charitable enough, and H&M's story is much richer than that: It involves uniquely human aspects of language, with humans enacting or bringing forth designer environments of public symbols, shared attention and common goals. This defense could continue, pointing out that natural origins of content is a properly embodied, extended and enactive account, which makes it significantly more powerful than disembodied evolutionary game theory. But is it? Not according to Evan Thompson:

> Beyond stating this proposal [of the Natural Origins of Content — T.K.], however, [H&M — T.K.] do not elaborate and support it with descriptions, analyses, or explanatory models of cognitive phenomena involving social learning, social cognition, and language. Instead, they switch to fending off critics [Hutto & Myin, 2017, p. 140 — T.K.] Although they claim to be doing naturalistic philosophy and deplore "the general tendency of philosophers—especially those in some wings of the analytic tradition—to assume that the essence of phenomena can be investigated independently of science" [Hutto & Myin, 2017, p. 276 — T.K.]—they do not draw from the rich cognitive science literature on how sociocultural practices and public symbol systems configure cognition. (I have in mind work by Lev Vygotsky, Merlin Donald, and Edwin Hutchins.) (Thompson, 2018, citations edited for consistency)

The latter point can be extended by pointing out that the body of important work on the origins of language does not eschew semantic content from the repertoire of their *explanantia* (Rączaszek-Leonardi, Nomikou, Rohlfing, & Deacon, 2018). These approaches seem to be making more progress than radical enactivism that Thompson accuses of being "enactive mostly in name only" (Thompson, 2018).

6. Conclusions

There is a thread linking the two stories I have considered here: the bootstrap hell argument and natural origins of content. This thread is the idea of unsupervised learning, i.e., exploiting the statistical structure of the environment to learn a better representation for describing or harnessing the environment. H&M are afraid that unsupervised learning is metaphysically impossible, on pain of circularity. We are lucky this is not the case, because the possible world of bootstrap hell is an insufferable place indeed: it lacks modern Internet, mobile devices and other conveniences H&M must have used in writing *Evolving Enactivism*, all of which are powered (in one way or another) by unsupervised machine

learning. Luckily, no intelligent mind will evolve there to experience this hell, because adaptive behavior also requires unsupervised learning.

It's quite reassuring that a way out of bootstrap hell and the solution to the HPC are to be found in *Evolving Enactivism*. It is the authors' own characterization of the natural origins of content as a bootstrapping process: Agents are capable of sharing content in a sociocultural niche in the presence of other content users. As I argued, this happens quite early on, possibly coinciding with the origins of life on Earth, and certainly before the mechanisms for transmitting genetic information evolved. The only problem remaining is that some will disagree that speaking of the social and the cultural is admissible at this point in natural history. That it is may be the one truly radical idea of H&M, although they will probably not appreciate this commendation.

Acknowledgements

The author is grateful to Marcin Miłkowski, Michał Piekarski, and two anonymous reviewers for their helpful remarks on earlier versions of the manuscript. This research was funded by the Ministry of Science and Higher Education (Poland) research grant DI2015010945 as part of "Diamentowy Grant" program.

References

Adams, R. A., Huys, Q. J. M., & Roiser, J. P. (2015). Computational Psychiatry: Towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery & Psychiatry, 87*(1), 53–63. https://doi.org/10.1136/jnnp-2015-310737

Alksnis, N. (2015). A Dilemma or a Challenge? Assessing the All-star Team in a Wider Context. *Philosophia*, *43*(3), 669–685. https://doi.org/10.1007/s11406-015-9618-2

Barrett, L. F. (2018). *How emotions are made: The secret life of the brain*. London, UK: PAN Books.

Bergstrom, C. T., & Rosvall, M. (2011). The transmission sense of information. *Biology & Philosophy*, *26*(2), 159–176. https://doi.org/10.1007/s10539-009-9180-z

Bickhard, M. H. (2009). The interactivist model. *Synthese*, *166*(3), 547–591. https://doi.org/10.1007/s11229-008-9375-x

Bouchacourt, D., & Baroni, M. (2018). How agents see things: On visual representations in an emergent language game. In E. Riloff, D. Chiang, J. Hockenmaier, & J. Tsujii (Eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 981–985). Stroudsburg, PA: Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1119

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001

Dennett, D. C. (2017). *From bacteria to Bach and back: The evolution of minds*. New York, NY: W.W. Norton & Company.

Dretske, F. I. (1983). *Knowledge & the flow of information*. Cambridge, MA: MIT Press. https://doi.org/10.1017/S0140525X00014631

Grice, P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388. https://doi.org/10.2307/2182440

Harvey, M. I. (2015). Content in languaging: Why radical enactivism is incompatible with representational theories of language. *Language Sciences*, *48*, 90–129. https://doi.org/10.1016/j.langsci.2014.12.004

Hohwy, J. (2013). *The Predictive Mind*. https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Hutto, D. D. (2018). Getting into predictive processing's great guessing game: Bootstrap heaven or hell? *Synthese*, *195*(6), 2445–2458. https://doi.org/10.1007/s11229-017-1385-0

Hutto, D. D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds without Content*. https://doi.org/10.7551/mitpress/9780262018548.001.0001

Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262036115.001.0001

Isaac, A. M. C. (2019). The Semantics Latent in Shannon Information. *The British Journal for the Philosophy of Science*, *70*(1), 103–125. https://doi.org/10.1093/bjps/axx029

Jelinek, F., & Mercer, R. (1980). Interpolated estimation of Markov source parameters from sparse data. In E. S. Gelsema & L. N. Kanal (Eds.), *Pattern Recognition in Practice. Proc. Workshop Amsterdam, May 1980* (pp. 381–397, 401). Amsterdam, the Netherlands: North-Holland Pub. Co.

Kolchinsky, A., & Wolpert, D. H. (2018). Semantic information, autonomous agency and non-equilibrium statistical physics. *Interface Focus*, *8*(6), 20180041. https://doi.org/10.1098/rsfs.2018.0041

Korbak, T. (2015). Scaffolded Minds And The Evolution Of Content In Signaling Pathways. *Studies in Logic, Grammar and Rhetoric*, *41*(1), 89–103. https://doi.org/10.1515/slgr-2015-0022

Lewis, D. (1969). *Convention: A Philosophical Study.* Cambirdge, MA: Harvard University Press.

Martinez, M. (2019). Representations are Rate-Distortion Sweet Spots. *Philosophy of Science*. Advance online publication. https://doi.org/10.1086/705493

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 3111–3119). San Diego, CA: Neural Information Processing Systems Foundation. Retrieved from http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf

Miłkowski, M. (in press). *Thinking about Semantic Information*.

Miłkowski, M. (2015). The Hard Problem Of Content: Solved (Long Ago). *Studies in Logic, Grammar and Rhetoric*, *41*(1), 73–88. https://doi.org/10.1515/slgr-2015-0021

Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=49593

Pattee, H. H., & Rączaszek-Leonardi, J. (2012). *Laws, language and life: Howard Pattee's classic papers on the physics of symbols with contemporary commentary*. Dordrecht, the Netherlands: Springer. https://doi.org/10.1007/978-94-007-5161-3

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York, NY: Basic Books.

Pelevina, M., Arefiev, N., Biemann, C., & Panchenko, A. (2016). Making Sense of Word Embeddings. In P. Blunsom, K. Cho, S. Cohen, E. Grefenstette, K. M. Hermann, L. Rimell, J. Weston, & S. W. Yih (Eds.), *Proceedings of the 1st Workshop on Representation Learning for NLP* (pp. 174–183). Stroudsburg, PA: Association for Computational Linguistics. https://doi.org/10.18653/v1/W16-1620

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. In M. Walker, H. Ji, & A. Stent (Eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 2227–2237). Stroudsburg, PA: Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1202

Rączaszek-Leonardi, J., Nomikou, I., Rohlfing, K. J., & Deacon, T. W. (2018). Language Development From an Ecological Perspective: Ecologically Valid Ways to Abstract Symbols. *Ecological Psychology*, *30*(1), 39–73. https://doi.org/10.1080/10407413.2017.1410387

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. Advance online publication. Retrieved from https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Rathkopf, C. (2017). What Kind of Information is Brain Information? *Topoi*, 1–8. https://doi.org/10.1007/s11245-017-9512-6

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00395

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shea, N., Godfrey-Smith, P., & Cao, R. (2017). Content in Simple Signalling Systems. *The British Journal for the Philosophy of Science, 64*(4), 1009–1035. https://doi.org/10.1093/bjps/axw036

Skyrms, B. (2010). *Signals: Evolution, learning, & information*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199580828.001.0001

Tani, J. (2017). *Exploring robotic minds: Actions, symbols, and consciousness as self-organizing dynamic phenomena*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190281069.001.0001

Thompson, E. (2018). Evolving Enactivism: Basic Minds Meet Content. *Notre Dame Philosophical Reviews*, *2018*. Retrieved from https://ndpr.nd.edu/news/evolving-enactivism-basic-minds-meet-content/

Wiese, W., & Metzinger, T. (2017). Vanilla PP for Philosophers: A Primer on Predictive Processing. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing* (Vol. 1, pp. 1-18). Frankfurt am Main, Germany: MIND Group. https://doi.org/10.15502/9783958573024

Yogatama, D., d'Autume, C. de M., Connor, J., Kocisky, T., Chrzanowski, M., Kong, L., … Blunsom, P. (2019). Learning and Evaluating General Linguistic Intelligence. *ArXiv Preprint, 2019*. ArXiv:1901.11373.

This publication has been financed by the Ministry of Science and Higher Education from the funds for the dissemination of research (DUN) within the framework of publishing activity, contract no. 711/P-DUN/2019, period of implementation: the years 2019–2020.