Representing As Compressing.

Bridging the Biological and Artificial Perceptual Systems: A Case for Information-based Explanations in Cognitive Science.

[DRAFT – DO NOT CITE, PLEASE]

Wojciech Mamak,
Section for Logic and Cognitive Science
Institute of Philosophy and Socilogy, PAS

Abstract

Psychology, statistics, computer science, machine learning, philosophy and many other disciplines routinely use information-talk in regard to a whole array of both biological and artificial structures. Thus, it is generally uncontroversial to hold that regardless of the differences and peculiarities of given cognitive systems, they are all information-processors. However, surprisingly, the notion itself (especially in the neural context) for long has escaped thorough philosophical exploration, as the other core concepts, such as representation (e.g. Chemero 2009, Ramsey 2007, Cummins 1996) or computation (Piccinini 2017, Fresco 2014) are frequently subject to. In order to do that I rely ideas from both information theory and philosophy, such as Daniel Wilkenfeld's recent 'understanding-as-compression' theory that frames compression-likes processes as maximizing scaffolding-to-output ratio. My aim is to show that the highly-structured nature of the informational makeup of the environment in which perception is performed (as demonstrated by the Natural Scence Statistics research program), it makes intuitive sense to think of perceptual processes as compression-like. Compression constitutes canonical, well-understood and elegantly formally defined information-

processing (Shannon 1948, Chaitin 1977) which could help explain how the perceptual system generalizes and generates information. Later on I assess the potential objections about the inability of information-based explanations to account for representations. To dismiss them, I turn to the recent theories of the naturalization of informational content (Isaac 2019, Skyrms 2010). I argue that it opens up the possibility of comparing artificial and biological signalling systems in the virtue of representational similarity. I mobilize a host of empirical evidence from both experimental neuroscience and artificial neural systems in order to assess how well the latter might model or explain the former. I also investigate how structural representational similarities between different cognitive agents can be traced due to the relation of infomorphism.

Keywords: compression, information-processing, perception, philosophy of cognitive science

The structure of the paper is as follows: in Section 1 I discuss some quite general principles connected to compression-like proceses that fit nicely with what we should reasonably expect of our theories of perception to explain. In Section 2, I offer empirical evidence from wide range of cases that demonstrates the robust presence of compression-like mechanisms in perceptual machinery of human and artificial agents and on different levels of processing. In Section 3, I possible worries about the fuzzy conceptual status of the notion of information, which could endanger the promise for its explanatory capacities. Then I shift to analyze the consequences of looking more deeply at the nature of perceptual representation as information-processing. Finally, I conclude in Section 4 with speculation on whether this strategy might be helpful in assessing structural similarities in terms of infomorphisms.

1. Regularities, compression and efficient perceptual coding.

'Brains are the ultimate compression and communication systems' – David MacKay declares at the beginning of his monumental ': *Information Theory, Inference, and Learning Algorithms* (MacKay 2003:V), a seminal book for computational modeling of cognitive processes Chaitin (2006) thinks that ‘compression is comprehension’, while Schmidhuber (2009) develops the idea of data-compression as basis for prediction in cognition. Compression-based explanation are found in neuroscience (e.g. Allen-Zhu et al. 2014), psychology (Wolff 1993) and recently – in machine learning (Goodfellow et al. 2016). It seems plausible to assume that there might be a philosophically-interesting connection between these quite diverse applications and perhaps some wiggle room for identifying structural similarities in the virtue of convergent information-processing.

Although quite a rare occurrence, philosophers do sometimes express interest in compression. For example, recently Daniel Wilkenfeld (2019) put forward a theory of 'understanding as compression'. He formalizes this idea as follows:

'*Understanding as Compression (UC):*

A person p1 understands object o in context C more than another person p2 in C to the extent that p1 has a representation/process pair that can generate more information of a kind that is useful in C about o (including at least some higher order information about which information is relevant in C) from an accurate more minimal description length.' (Wilkenfeld 2019:5)

Obviously this definition pertains to high-order, personal level of cognition, as one would reasonably expect. Wilkenfeld is quick to admit that as he states he is concerned with what understanding consists of for persons and concedes that might not be directly applicable to the types of compression we are most used to – artificial algorithms that usually deal with arbitrarily long strings of digits (or other mathematical representations) that are usually intractable for humans. Hence, it might be surprising that there is much to gain by tying compressing processes with human perceptual powers. But the general idea that motivates Wilkenfeld's account is a significant one.

He claims that both compression and understanding share their generative nature. To take his point across, he gives an example of two people – novice and an expert trying to prove a particular theorem – a soundness proof for first-order system. A novice uses a the most mechanical memorization strategy (with the Sheffer stroke operator [NAND], the only autarkic[1] operator), whereas the expert uses generalized knowledge of the matter to generate a full proof from simpler internal representations. Wilkenfeld notes that if any of the two provers happen to miss a step, they find themselves in wildly different situations. The novice faces a daunting predicament, as he is unable to notice and fill in the gap. The expert, however, is able to comfortably do that by generating missing piece of information from simpler representations. By invoking his definition, the expert understands better as she is able to generate more contextually-relevant information from minimal description (generalized knowledge of the problems contrary to the full-blown memorization). The invoked idea of compressing up to the 'minimal description length' is directly borrowed from Grünwald (2007), which neatly captures the intuitive idea that 'the more we are able to compress the data, the more we have learned about the data.' and directly nods to the algorithmic interpretation of information theory (Chaitin 1977).

Wilkenfeld's story demonstrates at least two of the properties that qualify compression-like processes as promising exploratory models for a host of cognitive tasks that we reasonably expect to require adherence to the same goals – generalization and generativity (Buckner 2018). According to Wilkenfeld: 'being in such a state where you have a representational kernel and an ability to piece together the rest is not just a result of understanding, but what it means to understand' (Wilkenfeld 2019:2). However, this idea can be understood more generally than just the feature of highest-level cognition (as understanding *prima facie* is) as pertaining to the cognitive ability of generalization (compressing[2]) and subsequent generativity (decompressing). To borrow his term, we can say that the idea of compression-like learning can be construed as a maximization of 'output-to-scaffolding ratio' (think again of the season logician example, which outperforms the meager memo riser in the virtue of having been able to produce any potentially missing step).

1 In a sense that all three elements of the functionally complete set {AND, OR, NOT} can be construed using only [NAND]
2 Here in a more restricted sense – as a part of the entire procedure of compressing/decompressing pair that forms the entire schema of 'compression (in a broader sense).

The idea of possessing an adequate representational kernel strikes as similar to the semantic pointer theory (Eliasmith 2003), which also explicitly refers to the notion of compression as an important feature of cognitive information-processing. Blouw et al. (2015), for instance, define semantic pointers as: 'symbol-like representations that result from the compression and recursive binding of perceptual, lexical, and motor representations, effectively integrating traditional connectionist and symbolic approaches'.

Semantic pointers are supposed to formalize (Eliasmith 2011) this idea of generalizing-generative function of compression. They are generalizing in the sense that they i) point to the perceptual information stored at the lower level (say: the images of flat surfaces with legs, the concatenation of letters 't-a-b-l-e', tactile sensations of tables, etc.); ii) retain semantic information extracted on the lower level. The generativity is supposed to come in the form of their readiness to be used to reuse the learned, compressed representation without the need to activate the perceptual-level states that the pointer 'points to'. This graphic summarizes the idea with a toy example:

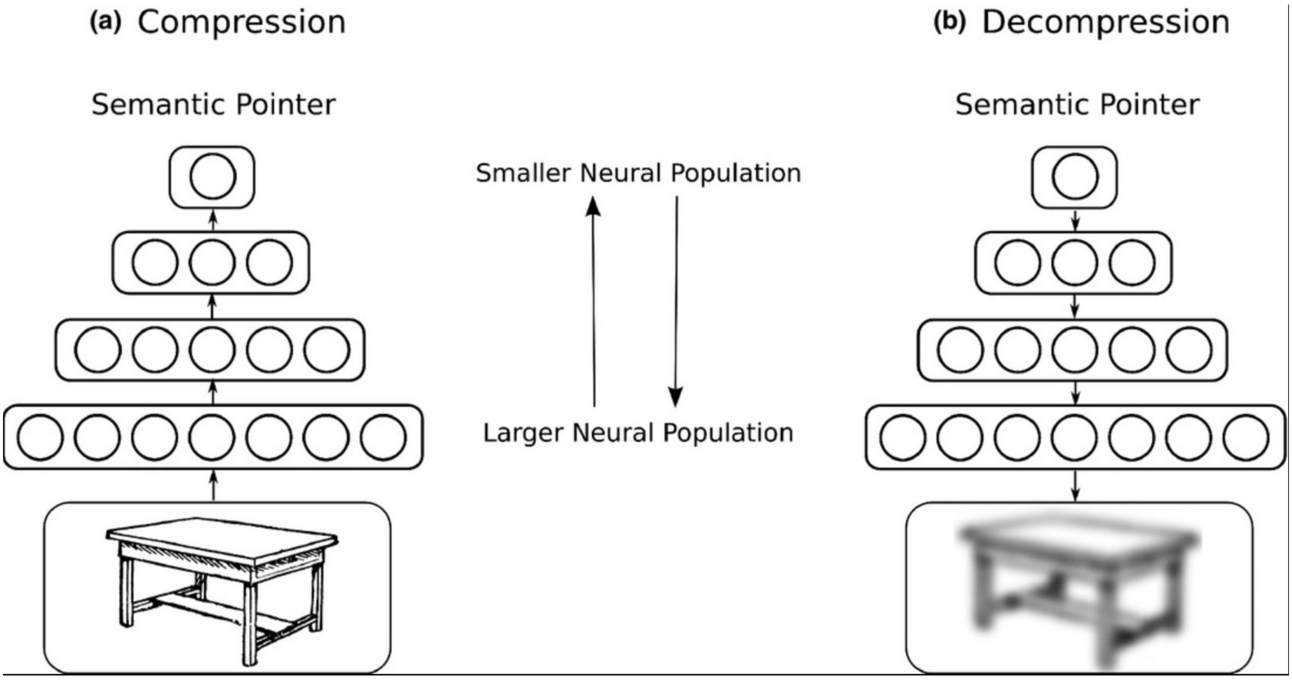

Fig. 1 – the schematic of the semantic pointer of TABLE. Multimodal information extracted and processed by the larger neuronal population are gradually compressed to form a more minimal representation. The subsequent evocation of the representation for some other task (e.g. recall) is achieved by the somewhat reversible procedure (compression is lossy, hence the decrease of the resolution'; adapted from Blouw et al. 2015)

A similar idea of looking for learned, *compressed* representations in information-processing perceptual tasks is pursued in self-supervised learning subfield within machine learning. For example Hjelm et al. (2018) suggest that the technique called achieves compressed representations by the maximization of mutual information between the encoder's input and output. This talk of representation-learning and representation-compression is commonplace in the subfield[3] (Poldrack, in press). Other information-theoretic measures, such as Kullback-Leibler divergence are also frequently used to explain the performance of the given network.

We see here that contrary to the semantic pointer framework (and corresponding SPAUN architecture), which invokes the semantic interpretation of 'information', machine learning techniques directly refer to classical sense of information. This apparent lack of clear definition of the term 'information' is a potential philosophical problem, because it might imply that there is a latent equivocation in place.

Neural contexts are especially prone to the usage of information-talk in quite a willy-nilly way, as a sort of intuitively-obvious 'semantic analogue of matter', which is simultaneously hidden in the environment to be mined from and then flowing (somewhat liquid-like) within the brain. Such a practice might not necessarily be detrimental if was restricted to informal contexts. All trades use jargon to communicate more freely and elastically after all and neurowizardry is a trade like any other. Unfortunately, the casual usage permeates scientific communication, which significantly obfuscates the non-trivial nature of neural information-processing (Baggio 2017). But it is conceptually dubious[4] to imply semantic function while at the same time one holds that it is an easily operationable (measurable and spatio-temporally tractable) quantity, as neuroscientist routinely do (e.g. Blinowska et al. 2004; Borst & Theunissen 1999).

3 An important caveat is in place here: both the data and the model can be compressed. The former can be conceptualised as the goal of processing, the latter is used to speed up the inference time (Goodfellow et al. 2016:448).

4 I am obviously not claiming that neuroscientists use the term incorrectly. Although it seems clear to me that a thorough conceptual analysis of the term to scan for possible equivocation would do the explanatory purchase of information-talk in neuroscience a world of good.

This is powerfully and poignantly criticized by Rathkopf in terms of informational objectivity and receiver-relativity (Rathkopf 2017, Rathkopf 2018). I do not intend to dwell on the subject too much, but Rathkopf essentially points out that any kind of information-flow cannot be 'taken out' as for instance blood flow does during the extracorporeal membrane oxygenation (ECMO) procedure. So the metaphor of actual flow collapses. But the task of disentangling this conceptual knot and defusing of the 'hydraulic metaphor' is indeed a huge task and well beyond the scope of this paper. The correct interpretation of 'information-talk'[5], especially in problematic neural context (Timme & Lapish 2018) is certainly crucial for the explanatory status of compression-like processes for perception (for example if it turns out that neural systems circulate different kind of information than artificial nets and computers do[6] or perform different kind of non-digital computation; Baggio 2017; Piccinini 2016), but for now we can simply agree with Griffiths (2001) that information in biology is still very much 'a metaphor in search for theory'.

For now, if ought to suffice to state that compression is usually defined as an information-processing procedure in the classical sense of the term (Shannonian) and described with tools and concepts from Information Theory or by its extension that is Kolmogorov-Solomonoff-Chaitin's algorithmic interpretation The simplest definition states that it is a process that encodes the original piece of information (signal) with the shorter string (lower bit-rate) than the original formulation. This kind of operation can only succeed when there is an element of statistical redundancy in the signal to be encoded. Redundancy can appear only if some kind of regularity organized the given signal or dataset. Entirely random processes are then fundamentally uncompressible. This crucial relation served Chaitin (1977) to define the very notion of randomness in terms of the compressibility of its respective signal.

5 The problem is even more complex. There does not seem to be common understanding on what constitues the vehicle for information carriage with the neural system. Most popular, Hebbian theory claim that spike trains are the answer, but other theories hold that the synchronisation is key and the bigger ensembles are the units of encoding (Baggio 2017). A sub-neuronal theory also starts to gain traction, trying to look for computations as deep as the dendritic spines). Also, as the notion of signal gets problematic, so is the noise. Accordingly, what was generally so far considered neuronal 'noise' finds its own evidence to play some part of encoding, too – especially in respect the motor functions (e.g. Musall et al. 2019).

6 The general problem is a philosophically old one – how to reconcile the mathematical, syntactic interpretation of information with all its formal elegance, breadth of well-proven tool and quantifiable, objective character with the intuitive notion of semantic information that seems to involve content and receiver-relativity. There are also other less pronounced conceptualisations of 'information', such as Bergstrom&Roswell's (2011) 'transmission sense of information', Godfrey-Smith&Sterelny's 'biological information' (2007) or Colgate&Ziock's teleological conception.

Put plainly, if there is no possibility of representing the original message in a shorter form (or in other words to generate a shorter-than-original program code), then it is entirely random (Dennett 1991).

Compression comes in either lossless or lossy form (Cover & Thomas 2006). The main property of lossless compression is its reversibility. The product of such compression, i.e. transformed (shortened) version of the original message can be inversely brought back to the original form without any piece of information lost in the process. Lossy compression, on the other hand, is a trade-off procedure. Some information is lost in the transformation, so the process is not fully reversible and the relata of the transformation are not mathematically identical. But the payoff in terms of the saved storage (shortened length) is usually much greater than in the lossless-type procedures (hence it's a trade-off). Hence, it is most likely to think that human perceptual system mostly in concerned with lossy compression. Bates and Jacobs (2020), for example, claim that perceptual processes instantiate coding-procedures defined formally by the 'rate-distortion theory', which is concerned with efficient communication with the minimization of loss of information during the encoding/decoding procedures.

But how does this elegant, yet abstract, mathematical set of ideas correspond to perceptual phenomena? The most trivial observation is that both deal with selective, regularized, generalizing information-processing, so there is *some* correspondence But that's surely not sufficient. More importantly, though, perception would not succeed in a random environment. The goal to explain perception is to explain how it operates in a highly structured, regularized environment. This regularization points at the notion of predictability. As we saw earlier, the predictability of both the stimuli and the signal is of crucial importance for efficient use of regularities underlying the informational landscape the agent operates in. That prima facie suggests that compression-like usage of the regularities can help explain the basis of the important feature of perceptual systems that is their ability to 'fill in' the details, as in numerous example of perceptual completion in psychology (Kanizsa 1985; Sekuler & Palmer 1999). It can also serve as a rival model for furthering the debate on the use of purported 'knowledge' in perceptual processes. The relation I have in mind can be spelled out in a simple maxim that links perception and compression on theoretical level:

Redundancy = regularity = predictability.

I will build on this rather general set of observations and try to analyze how they contribute to understanding perceptual processes as efficient coding procedures that are compression-like. For the sake of brevity, I restrict the scope of interest to the visual tract.

Perception is a heavily resource-limited enterprise and takes advantage of that principle. Sensory detectors are attuned only to particular classes of signal. Their resolution varies greatly within the system, too. For example, visual acuity is much greater within the fovea (where the density of rods and cones is much higher than in the peripheries), rendering the on-center (of fixation) sight much sharper than off-center (Tootell et al. 1988). But that's just the start, as just a step later the visual processing encounters the first obstacle in the form of bandwidth of the optic nerve that relays information to the cortical structures (or rather more precisely – first to lateral geniculate nuclei in the thalamus that is only later imported up to the cortex). The throughput that the optic nerve can transmit is much smaller than the capabilities of the population of photoreceptors. Their output must by then compressed in some way. The process that is responsible for that is called whitening (Field 1987) and is performed in retinal ganglion cells, just before the signal is sent downstream the aforementioned chokepoint (informational bottleneck; for the computational discussion of the idea, see: Tishby and Zaslavsky 2015). High-spatial frequencies (less common in the environment, responsible for coding abrupt changes, i.e. edges) are weighted higher than low-spatial frequencies (coding smoother transitions in the images) receive less weight and are thus effectively filtered out.

At the cortical level, this anatomical limitation is virtually absent. Neurons in neo-cortex are highly interconnected, giving more degrees of freedom to implement different ways of efficient coding. Sparse coding is one of them. In fact, different representations at the neural population could all decode the retinal input (Foldiak & Young 1988). From the very extreme of an dense code, with only a few neurons, all being somewhat active at any given input, to a very local code with many neurons but only one being active at any given input. Sparse coding is in the middle of these two extremes. But, depending on the definition of efficiency, all three codes can be considered efficient (depending on the combined cost of having many neurons, the cost of having many interconnections between neurons, metabolic firing cost[7], etc.).

Whitening is not a necessary component of sparse coding. These are two different efficient coding principles and compression procedures that seem to co-exist at different processing stages in the human brain. A few side notes here. It may well be possible that these two coding strategies are formally or mathematically related, and that both can be reduced to the same learning algorithm. This search for a single underlying principle is an active topic of research for theoretical neuroscientists. Recent advances in deep nets, or convolutional neural networks show that some architectures automatically develop a sparse code, without a sparseness constraint being explicitly built-in (Li et al. 2013). In computer vision, pre-whitening of the input leads to faster learning of a sparse basis set of components (and also a less noisy basis set, Le et al. 2011). So, although whitening is not a necessary component of sparse coding, it seems to improve the results of a sparse coding algorithm[8].

We will come back to the point about the striking similarities between the findings in computational neuroscience and in engineering artificial neural nets. These kind of principled, information-based relationships suggest that the similarities are non-coincidental. That suggests that understanding artificial nets as models can truly help us to further our understanding of perceptual processes.

Please note that we can differentiate between plain insensitivity to the

7 The energetic cost is a truly important one for the biological systems. Brain are fantastically efficient in that respect, utilising ~12W (which still constitutes a whooping 20% of the organism's total energetic intake) (Żarnowski, forthcoming), much more efficient that any current artifical systems would achieve. It is because in computer science it is usually time and memory that are the resources that need to be optimize. Energy is assumed to be free real estate, the least important factor of any trade-off. This is important to bear in mind while tempted to assume that artificial systems can be optimised according to the same constraints on purely computational terms.

8 It might be worth to point out that we are then involved with the algorithmic level, speaking in Marrian terms (1982; Shagrir & Bechtel 2015).

particular parts of the environmental signal and actual filtering. Human eyes are sensitive only the a relatively narrow spectrum of the electromagnetic spectrum (the 'visible spectrum' or the 'visual window'). We are not inclined to say that we filter out the, for instance, infra-red light in any sense, but rather not equipped to handle this particular type of signal. Contrary, low-spatial frequencies (within the visual window) are actively suppressed. Certainly, the neural signal-processing picture is much more complicated and includes also non-linear procedures such as threshold and saturate functions, pooling, normalization (DiCarlo 2012, Carandini, 2003, Lennie & Movshon 2005), etc. but for our present purposes it suffices to say that there is an important difference between certain 'deafness' (or insensitivity) to particular (potentially informative, hence behaviorally advantageous, after all) signals and their active filtering. Only the latter can be meaningfully described as compression-like and only the latter is a perceptual activity, as is caused by the stimuli actually impinging on sensory apparatus.

We can think about this problem from the different perspective when we take into account how important is the format of informational intake. Seeing a particular image, say, a portrait usually does not pose too great of a challenge to the human perceiver. Let's suppose, for the sake of clarity, the image is represented in grayscale of low granularity (so there is just a handful of particular shades of gray, say 10: coded from 0 to 9). The picture is still quite easily processable and perhaps even recognizable (if the face is familiar). Now, if we 'translate' the grayscale values to their corresponding numerical values displayed as actual symbols (so every pixel on the matrix would be substituted by its respective greyscale value; e.g. All the black pixels would be substituted by zeroes, white pixels with 9s and all the intermediate levels with respective digits). Even though that representation is still mathematically equivalent (structure-preserving, homomorphic), it ceases to be readily available (Hyvaerinen, Hoyer, Hurri, 2009:5).

A similar idea was discussed by Dennett (1991:33-34) when he considers translation of the 900-pixel bitmap into 4-bit chunks and then substituting them with equivalent hexadecimal encoding. After such a transformation the previously easily recognizable patterns in the bitmap (and the potential intrusion of noise) becomes intractable in the hexadecimal string.

We can see how even simple static-image processing can be potentially problematic for perception. But real-world visual perception is perhaps unsurprisingly, an infinitely more complicated task. If we think of the environment as the dataset, it is as high-dimensional a set, as they come. Thus, any cognitive agent that challenges it – be it a paramecium, a robot, an animal or a human being faces the same reality, under the enchantment of what is sometimes called in computer science 'the curse of dimensionality' (Bellman 1956). It can be boiled down to the idea that as the number of dimensions grows, the number of samples needed to theoretically cover the entire space of possible inputs grows exponentially[9]. Any successful cognitive systems has to utilize representations, because the independent 'one-to-one' extraction of every single feature of the environment in isolation is computationally impossible (Poldrack, forthcoming).

It really seems to be a daunting challenge. Nevertheless, somehow all the organisms that inhabit the 'world out there' face this very curse yet do not seem to care and cope just fine. But what is perhaps even more interesting – recently the machines started to escape the curse, too. Hinton and Salakhutdinov (2006) in a classic paper demonstrated how the neural networks perform this kind of dimensionality reduction (sometimes called perhaps more technically and certainly more dramatically – the dimensionality collapse). They achieved that by mapping the statistical regularities of the environment (dataset) onto initial weights of the intermediate layers (through pretraining). Although the technical side of that should not concern us that much, it is important to understand that we can now plausibly demonstrate how the usage of structure in the environment (regularities in the dataset) effectively constraints quasi-perceptual performance (although admittedly in isolated tasks such as image classification, so far).

9It is perhaps helpful to think of it as an analogue problem to the frame problem, well-known in philosophy. Although frame problem is concerned with first-order logic, whereas curse dimensionality describes the problems of data-analysis, there seems to converge on the signalling the difficulty in constraining relevant knowledge to effectively navigate an exploding space of possibilities. It is a problem of how to decide which features are relevant without the prior analysis of them all (which is obviously impossible because of their gigantic volume).

The generalization of the machine learning framework for perception is still a huge 'if', but some crucial pieces of evidence are already in place (and rapidly accumulate). That is why it is more useful to talk not of wholesale perceptual optimization strategy, but a combination thereof. These strategies are relative to both the stage of the perceptual processing (as in case of whitening and sparse coding), but also particular task. It is important to remember that conditions (constraints) on the optimality of the process are relative to how we delineate the system and assess its goals. This might seem trivial at first, but should make us vary of speaking of global optimality in perceptual processing all along the (e.g. visual) stream and also incline us to analyze the compression-like powers of the particular parts of perceptual systems, respective to their functional parts, contrary to the ambitions of the likes of Wolff (1982, 1993).

To summarise, we can invoke the three principles for efficient compression in perceptual systems as recently proposed by aforementioned Bates & Jacobs (2020:3-5). Their formulation seems to recapitulate nicely the properties we discussed so far in favor for understanding perceptual processing through compression lens. These three principles are in their terminology:

Limited capacity principle: which states that due to cognitive limitations of human subjects as perceivers, they are doomed to apply the generalize-then-generate kind of procedure in order to cope with the highly complex, yet structured environment.

Prior knowledge principle: which states that to compress efficiently a knowledge about the statistical properties of the signal are necessary. Their formulation in terms of actual knowledge (such a view is outspokenly defended also by Palmer [1999]) is, of course, problematic (and as I argue in later section, while discussing the representational format of information – misleading and unnecessary). But regardless of whether we treat the statistical model as internally represented or not, or rather it is 'knowledge-within-the-system' or rather 'knowledge-to-the-system', we can define this condition as taking advantage of regularities in the signal while explaining the systemacity of perception.

Task-dependency principle: which states that any compression-like procedure to be effective with respect to a certain behavioral goal of the agent.

Now I would like to turn to some more empirical evidence of such specialized compression-like strategies being employed by the visual system.

2. Empirical evidence

We have so far briefly discussed the simple case of static image-processing and mentioned that the visual landscape is significantly more complex. The brain copes by utilizing the deep regularity ingrained in the complex, yet systematic environment, as evidenced by the Natural Scene Statistics paradigm. We know that pixel intensities in the pictures are not randomly, but highly-autocorrelated (Machilsen 2014), the same goes for edge-orientation (Geisler 2001) and even perhaps categorical features (Torralba & Oliva 2002). But the picture (literally and metaphorically) gets blurrier, the faster the image moves.

This intuition is widely utilized in video-processing community. Moving picture, contrary to static signal, presents an opportunity for two complementary compression strategies – intra-frame compression and inter-frame compression. Intra-frame compression equals roughly treating the particular frame as a static image and processing it according to static-image compression algorithms (e.g. PNG). But the inter-frame coding gives opportunity for more greedy approach. It makes intuitive sense that representing every next frame literally, in a sort of bitmap seriatim processing, is hugely inefficient. After all, we can group the block of frames, assume they would not differ that much (similarly to the assumption of autocorrelation of pixels in static images – we take advantage of the fact that natural world with disregard to all appearance of turbulence, is really quite stable). If that is the case, we needn't represent every new frame in isolation (and in a sense, from the scratch), but can only register the difference between the given one and its predecessors. This is why livestream videos containing snowfall or confetti dramatically lose resolution, whereas the static frames with them are perfectly sharp. The video processors do not work like that, they try to computer the difference (e.g. to perform a simple transformation of form move this batch of pixels by this many pixels) or sometimes use simple memory storage (e.g. MPEG-2 format does). Of course I am simplifying here greatly – the video coding is an extremely complex craft and the abundance of techniques is a direct corollary of their variety. Nonetheless, we can be content with the simple observation that the capture of moving pictures creates new dimension for structure, regularity and predictability, which is enabled by movement. Hence, we can reasonably ask – does our perceptional system take advantage of that and compresses accordingly, too?

This problem was investigated by the research group led by Nortmann (et al. 2013), which analyzed the reaction of feline primary visual cortex to such rapid changes of stimuli through voltage-sensitive dye imaging. Researchers presented the subjects with sequences of images (natural scenes) and their superpositions with edges of particular orientations removed. The results were surprising. According to them, V1 actually employs two different coding strategies, depending on the time elapsed between the differing versions of the image. On a shorter timescale (30 milliseconds), the neurons represented complete image information, dubbed the 'ongoing representation'. But as the elapsed time between the sequences was extended to more than 100 milliseconds, the neurons represented only the image differences (what was new or missing, depending on the order of presentation). Such evidence for dual computation purely on such a very early level is surprising Visual cortex suppresses redundant information. By doing that it saves resources by frequently forwarding image differences, pretty much as the methods used for video data compression we discussed earlier.

It is important to notice here that both in the visual compression case and in the Nortmann's study it is very tempting to say that what explains the process is some kind of prediction. Indeed, in both cases that is exactly the type of language we encounter in their analysis. This is routine in discussion of compression schema in video processing. For example, Mavlankar et al. (2006) when advocating for the particular coding scheme of their design, they claim that its 'intraframe prediction provides an alternative and increases coding efficiency'. The skeptic would be probably quick (and right) to object that this kind of talk is simply technical, disciplinary jargon of little philosophical interest, as everybody would agree that neither cameras, not the compression schema predict in any philosophically robust sense of the term, even if the user, for once, downloaded all the right codecs.

Let's move to the physiological case, then. The authors themselves admit predictive interpretation as the most plausible one. To quote the passage: „responses to stimulus differences fit well conceptually with predictive coding principles, proposing that deviations from cortically generated predictions are propagated up the visual hierarchy as error signals”. But that seems hasty at best. The same effect can be explained by adaptive behavior The bias inherent in the presented signal is counterbalanced by the visual system by introducing its not 'counter-bias' in responsiveness and selectivity to achieve so that the average firing rate of given neuronal population ('population signal') was kept constant. Such a principle has been long advocated by psychophysicists maintain equality in the time-averaged responses across the population and to enforce independence in selectivity across the population. The evidence for this homeostatic hypothesis has indeed been found by a team led by Benucci (Benucci et al. 2013).

The structure can be meaningfully described as performing information-compression as it actively suppresses redundant informational intake to be sent further downstream. It is an optimizing, efficient neural-informational procedure. But it's no more predictive than the optimizing, efficient information procedure of video compression. A spectre is haunting neuron – the spectre of predictivism

Compression seems to be a case of biased process defined over informational states. But how can informational states fare better in explaining perceptual processes? In the next section I briefly consider some objections to information-based account and suggest how these worries might be alleviated by existing theories of naturalization of informational content. I am mostly concerned with the problem of bridging information-based principles with accounting for representational powers in perception. I argue that holding firm on the informational analysis opens the gate for meaningful explanatory mapping from artificial to natural information-processors that keeps their representational capacities, but does not overintellectualise the problem like commitally predictive or inferential accounts do.

3. Informational states and representational capacities

The common charge against the information-based explanations accuses them of the inability to account for content. The informational states are supposed to be not about things as such unable to contribute to the meaning. The orthodox interpretation of information theory seems in line with this assertion. As famously put by Freeman Dyson to information-theory 'meaning is irrelevant' (Dyson 2011; per Shea et al. 2017:1009). Shannon himself was openly skeptical about the extensions of his theory to anything more than a mathematical theory of effective communication and formal description of signaling channels. He considered the ascription of meaning to information-theory to be a fundamental misuse of its core function and attacked it in article forthrightly titled 'The Bandwagon' (Shannon 1956).

The rudimentary problems seem to be how to account for the shift from simple signaling systems to contentful representative powers of the given systems. Precisely this riddle seems to motivate the constructive account of cognition and perception itself. Predictive processing, inferentialisms of all sorts, enactivists share the belief that the scaffolding has to come from above all the way on the ladder. In perception this problem is chiefly motivated by what is called the 'inverse problem'.

The old chestnut of 'inverse problem' in philosophy of perception is arguably as vintage as it gets within the discipline. The problem, which basically pertains to the troubles with explaining perceptual experience in the light of underdetermined stimuli, seems to come back in ever newer iterations from at least the Kantian times (Hatfield 2002). The riddle is, again famously, whether perception can be explained purely by the states of the world-out-there or rather some kind of agential interpretation is necessary for any kind of perceptual process to succeed. Obviously, the most important milestone was set by Helmholtz (1867) and his hugely influential doctrine of 'unconscious inferences' in perception that were supposed to solve the problem for good. Unsurprisingly the establishment of the Helmholtzian tradition and its long lineage (Gregory 1998, Marr 1982, Rock 1983) was only the onset of the much-prolonged trench warfare between proponents of inferentialism and its adversaries – the successors of the empiricist tradition that claimed that perception is a passive process of receiving particular kind of information about the environment, such as Gibsonians. This venerable old debate has found new fuel and exciting reconceptualisation with the advent of yet another round of active perception theory, such as Bayesian theory and predictive processing framework. Leaving aside the differences between the two, it suffices to say for now, that both claim perception is necessarily involved with generating rivaling hypotheses about the incoming perceptual input, against which the proximal stimuli are then judged against. These hypotheses (or assumptions) are encoded as priors and hyperpriors (the latter differ from the former by their wider scope or: being of more general nature; the information that the light is generally expected to come from above is hence an example of a hyperprior) and then used to calculate posterior probability of the percept (given the particular hypothesis) (Orlandi 2015).

A contrary view, representing the non-constructivist camp rejects the full-blown inferentialist view as over-intellectualizing perception. The rival view holds that there is a simpler, yet more empirically-grounded framework to explain perceptual processes that does not resort too much else then the properties of the perceptual signal itself to solve the inverse problem. This family of views is based on the Natural Scene Statistics (NSS; Hyvarinen et al. 2009) program in perceptual psychology and artificial intelligence that holds it is the thorough analysis of the properties of the signal itself that can explain away a lot of the inferentialist constructs. It was recently most eloquently and powerfully put forward by Orlandi (2014; 2015) and dubbed 'The Embedded View (EV)'. The EV basically holds that perceptual processes are better understood as biased processes defined over non-representational states. Orlandi however, does not discard the Bayesian outlook, but offers its major reinterpretation She claims that Bayesian inferences are not legitimate, but merely metaphorical inferences. If the inferentialist concedes that, however, his position loses any purchase as a constructivist theory, a real alternative to the ecological perception.

What is crucial to understand in this debate is that both sides agree on the fact that perception is a non-random process that heavily utilizes regularities, although they differ on the origins of these regularities. Both also seem to agree (as virtually all cognitive theorists do) that perceptual systems are in the information-processing business. When one takes both of these points in consideration, it becomes quite startling that a crucial idea about the usage of regularities in the informational input has been mostly neglected. That idea is compression. One of the main points of this paper is to argue that it is favorable to use this concept to better understand the biased nature of perception without the resort to vague and unproductive overintellectualisations. Hence, I agree with Orlandi with her general statement on the nature of early visual process and suggest a complementary way to argue for a non-constructive account.

Let's go back to the problem of emergence of content that purportedly disqualifies informational explanations for representational powers. The obstacle on the way is called 'the hard problem of content' (HPC), which posits that the criterion for being a contentful state is to possess satisfaction conditions. This argument has been recently refreshed by Myin & Hutto (2013; 2017)(in their widely discussed argument of 'bootstrap hell' against the contentfulness of basic cognitive systems. I do not want to dwell further on their proposal, suffice to say for now that according to advocates of HPC informational states are supposed to merely lawfully co-vary with the environmental features without the property of being about them. Think of number of non-decayed atoms of C14 carbon within an archaeological object (say, a jug) that lawfully co-varies with how old the sample is, but the ratio of decayed and non-decayed atoms itself is not about the age of the jug. There is information within the sample (in Shannonian) but there is no bridging that with the contentful, semantic account of information. The jug surely does not represent its own age. Fodorian case of rings on a tree is another classic example of a structure that lawfully co-varies with some property (age), but is not 'about' in a within-the-system sense (Fodor 1999).

I think this kind of objection is misguided and there is already a host of promising philosophical theories that suggest the way to circumvent the HPC or render it invalid. Skyrmsian theory of evolutionary game-theoretic signaling systems, building on Lewis' (1969) seminal work on signaling games is one. Shea, Godfrey-Smith and Cao (2017) also demonstrated how simple signaling systems can achieve states that satisfy the required condition of being contentful A slightly different, more conceptual approach of demonstrating that Shannonian information itself is indeed semantically-rich, has been recently put on the table by Isaac's work on latent Shannonian semantics (2019) and Martinez's (2019) idea that Shannon's information is equally a theory of communication and representation.

The problem was sometimes stated in terms of possibility of misrepresentation. Informational states are not contentful because lawful covariance does not leave possibility for the system to 'be wrong' about what it informs about. These kinds of worries were addressed by Dretske's protopredicative indicator theory (1982), Millikan's teleosemantics (1984) or the idea asymmetric-dependency criticized by Fodor (1990) by trying to offer a more minimal account of representation. However, the skeptics of minimal representation in simple cognitive/perceptual systems remain unshaken, pointing at other criteria that ought to be satisfied before we should be inclined to admit the representational status of a particular information-processor (note that label is virtually uncontroversial across the board).

One of that important criteria is decouplability. The ability to act on absent stimuli (informational intake) is yet another. A somewhat similar desideratum to the former is sometimes formulated as 'distance' from representation to what it co-varies with. Let's start with the inverse order.
Distance is conceptualized as the capacity of representations to occur (in principle) in the absence of their typical causes. This is supposed to

understrike the nature of informational states as mere causal mediators or 'relays'. States that (even asymmetrically) co-vary with something and/or have the function of doing so are *too close* to what they causally and statistically correlate with (Orlandi 2014:121).

Absence is another criterion. A handful example is provided by Cantwell-Smith (1996), who considers the example of coyote and rabbit. Rabbit carefully keeps on an eye on the coyote (as he is one clever beast). As far as the coyote is within the line of rabbit's sight, there is a connection between the predator and prey's perceptual system. However, when the coyote turns a corner and hides behind a rock, the previous link is broken, yet the rabbit still guides his behavior according to the position of the coyote. In a sense, now the rabbit deals with something that is perceptually absent to it. A corollary of the ability to represent absence is the ability to 'fill in' the holes in perceptual outlook. 'Completion is an important mark of the emergence of representational states. It indicates that perception allows us to track what is in the world even when what is in the world is, in some sense, not present.' (Orlandi 2014:126) Gestalt-like laws or constructive (inferential and/or predictive) accounts are usually mobilized to explain this important feature of successful perception. But as we have seen, compression-like systems that thrive on statistical regularity in pattern recognition are custom-made for dealing with such biased processes just as they are – just biased.

All three intuitions – absence, distance and completion can be generalized to the idea that representations has to be in some way disattached from the online perceptual input. Numerous authors (Clark 1997, Chemero 2009, Gallagher 2017) has indicated that this notion of *decouplability* plays an important role in explaining representational character of cognition.

This idea of detachment from 'online' stimulation seems to mirror a very old philosophical intuition that seems to get a bit of a bad rap (or at least a label of antiquated, folk concept). Namely the idea of cognition/perception boundary. It is precisely that independence (in a loose sense of the world, cognition is in almost all accounts *somehow* inextricably connected to proximal stimulation) from the actual sensory activation. There are of course complications to that account, such as endogenous hallucinations[10], demonstrative thoughts and perceptions of absence[11]. Beck (2015) defuses them nicely and offers a plausible definition of perceptual processes to differentiate them from cognitive ones[12].

According to different flavors of multi-layer predictive theories the existence of 'joint in nature' (Block 2014) between perception and cognition has been rendered obsolete by the new framework. There are supposed to be in fact many more interacting layers acting on the roughly similar organizational principles (say, hypotheses tokening and testing). But as we see – the notion of decouplability seems to be just another way of sneaking the old distinction to the debate through backdoor yet again.

The problem of conceptualizing the interface between perception is sometimes (most famously by Fodor) conceptualized as the 'information encapsulation problem'. For Fodor the stake was to uphold the modular view of the mind. He (along Pylyshyn; 1999) devised the idea of encapsulation to fend off the New Look psychology (spearheaded by Bruner and Postman; Lupyan 2016). But the debate eventually floundered with the inability of phenomenological (e.g. argument from stability of illusions) and behavioral evidence to account for definite encroachment of the modular borders (from higher to lower levels). A little semi-historical note is perhaps in order here that Fodor was extremely skeptical of the type of imaging evidence. Most famously, he lambasted the usage of 'coloured maps of brains' as tools for inferring the informational content of brains in his 1999 article for *London Review of Books* (Fodor 1999). Most recently Coltheart (2006) expressed similar worries that imaging evidence is next-to-useless for mapping the psychological phenomena into physiological ones. I argue we now possess

10 Think of acid-wrenched Johnny Depp getting assaulted in a bar by the reptilian envoys of the New World Order in *Las Vegas Parano*. Not to be confused with the thirsty exhausted wanderer in the middle of Saharan caravan who experiences a mirage of an oasis – an exogenous hallucination.

11 An important clarificatory note is in order here. Perceptions of absence are not to be confused with 'absence perceptions'. The latter are types of cases as discussed in the rabbit and coyote example, when the absence of something (coyote) is a part of bigger cognitive picture (the entire episode and the perception of the rest of environment, such as the rock). Perceptions of absence are situations when an agent realized there is no stimulation (of a given modality) whatsoever – such as when the sudden silence strikes.

12 His criterion for being perceptual is formulated as a 'full functional stimulus-dependence': „ψ is perceptual if, necessarily, all occurrences of all elements of ψ have the function of being causally sustained by present proximal stimulation; otherwise ψ is cognitive.

better tools to take the 'informational' part of the encapsulation problem more seriously and try to present novel techniques that display promising results for such an enterprise – artificial networks as model and sophisticated imaging techniques.

4. Structural and representational similarities. Infomorphisms?

But the newer techniques such as Representational Similarity Analysis (Kriegeskorte et al. 2018; Roskies, in press: for philosophical analysis) The seminal paper on the method (Kriegeskorte et al. 2008) offers a tour de force of the method. The aim of the study was to present the same set of visual images (the set contained many categories of objects) to humans and macaques and register the activation of the inferior temporal cortex (through fMRI in humans and electrophysiological recordings in monkeys. Then a clustering algorithm was used to build a matrix representing the response of recorded voxels (humans) or neurons (monkeys) in response to particular categories. The analysis of similarities of the matrices discovered a similar mapping that entailed some semantic-like relations between objects (living and non-living kept separate, followed by according subcategories – e.g. living creatures, then body parts) In each case, a clustering algorithm was applied to the response to a large number of simultaneously recorded neurons/voxels in order to visualize a low-dimensional embedding of neural responses to specific images. Across species there was a highly similar clustering of visual images, with a distinction between living and non-living objects at the top level, followed by subcategories within those larger categories (such as faces versus body parts, or natural scenes versus objects). That suggest RSA allows for cross-format structural comparisons and also gives hope that we can infer representational *contents* from proxy vehicles (recordings). Kriegeskorte do not hesitate to explicitly call the representational similarity project a 'shift from activation to information' (Kriegeskorte et al. 2008). Interestingly Bates and Jacobs report (2020) that the use of their compression-based model (called rate-distortion theory) that operationalizes compression as an organizing principle for training, representational space similar to reported to Kriegeskorte's studies, emerges.

I argue that to do that we need to be able to reliably track information flow (and its direction), but also account for the right format, i.e. identify according representational structures. For example, inferentialism holds only if the format is proposition-like. But that is difficult to empirically validate (which perhaps explains and persistence of these theories). Thus, it seems to me that Orlandi's explicit ambition to provide a conceptual framework to disentangle constructive from non-constructive processes in vision is well-motivated one. The difference hinges on the nature of regularities used by the perceptual system. Her idea is to introduce the differentiation between representations of sensations (*s-representations* in her terminology) and representations of principles (*p-representations*. The goal of that is to differentiate the system has the regularity encoded within itself (has the relevant p-representation) or merely behaves in accord with that rule in a trivial sense (as Mondays behave accordingly to the rule that they come after Sundays) (Ramsey 2007). This is an extremely important point for the success of genuine explanation of informational and representational structure of the system. We have to clearly delineate what is information-within-a-system to information-to-the-system.

The dichotomy comes in handy for example when analyzing imaging studies. When we hunt for voxels and find out they regularly co-vary with a particular task of stimuli, that information that the voxel is vehicle for may not necessarily be the vehicle (true information carrier) for the brain (Egan 2014). On the contrary, it almost surely is not the case – voxels are artifacts and do not map nicely onto structural boundaries between brain structure. They are also somewhat coarse-grained, as each voxel is a cube of 3 millimeters containing millions of neurons. Certainly, voxels are reliable indicators of information-processing within the brain so their activation do seem to co-vary in some (much) better-than-chance way with the brain activation. But it is unwise to assume that there is a strict one-to-one mapping or any kind of identity relation between the two.

This is precisely the guardedness that keeps us in line with parsimony and accuracy of our theory. If we can explain Mondays, digital cameras and perhaps V1 cortex ensembles – all of we know about them – without the resort of high-level predictive powers, then we should do without predictive language, same goes for inferential explanations (e.g. Kiefer 2017). We theoretically could treat them 'as-if' they predicted or inferred (as it is sometimes valuable – for example for didactic or heuristic purposes), but for explanatory purposes: why should we? That is not only confusing – implying deep connections between phenomena that are not necessarily related, but also explanatorily barrening. Overextending concepts trivializes them (if almost everything performs a certain function F, it no longer tells us anything that X Fs; as we discuss information-processing we can say that as the probability of X F-ing approaches 1, the informational value [surprisal] near zero). Point in case – the notion of inference was overextended precisely for the purpose of explaining the purported constructive aspect of perception (Lupyan 2016) or cognition more sweepingly (Friston et al. 2011). Despite the valid criticism of Orlandi (2016) that Bayesian inferences are not genuine inferences (operating on representational structures in according to language-like rules), the proponents of predictive account double down on the idea of literal interpretation perceptual inference (Kiefer 2017).

But it probably affords to object now that if inferences and prediction are deemed to be ascribed too liberally, why not be equally suspicious of representations. If we are advocating for informational account, should not we try to discard representations, too? This, however, assumes that representations are just theoretical posits (as they have been a central one in psychology since its inception). Interestingly, we might be pass the point where this is the case, as convincingly argued for by Thomson and Piccinini (2018). Neuroscientists do, in fact, routinely and comfortably not only observe, but manipulate them, rendering them perfectly real scientific objects (Hacking 1983)[13]. The neural representations come in different sub-classes. Piccinini and Thomson differentiate sensory, motor, and 'uncoupled' (equaling decoupled ones in the sense discussed earlier) representations. Such real neuronal representations are to be found even in early vision, where the standard example of sensory representation of absence is the blind spot (Matsumoto & Komatsu, 2005). Sensory representations seem to be in line

13 To be precise, according to the authors, this is by no means a novel phenomenon in neuroscience (they claim it to be at least a century and half old, reaching to Hughlings Jackson's experiments on motor representation). They simply claim that philosophy has been extremely sluggish to catch up.

with the desiderata for s-representations *sensu* Orlandi.

The worry about the overly liberal policing of what counts as representation is based on the prima facie plausible assumption that 'while information is ubiquitous, representation is not' (Orlandi 2014: 105). That should make us vary of equating informational processes with representational ones. Granted, not all information-processing systems are able to represent. Moreover, almost none that trade in information do reach the status of truly (cognitively) representational systems. But ever more plausibly, some do. And for sure – all truly representational systems do trade in informations, too. Question is – do they represent in the virtue of attaining a particular information-processing structure.

Quite famously by now, artificial systems such as deep neural nets rapidly achieve impressive success in a wide range of perceptual task, such as image categorization, image generation, classification, etc. The neck breaking tempo of proliferation of new architectures and techniques might imply the belief that in fact many diverse systems can realize a lot of perceptual functions. But that is misleading. Cao and Yamins (forthcoming) observe that in fact only simple tasks have many solutions. More complex tasks are by definition more constrained (have to satisfy many conditions at once and trade-off between them efficiently). The more constrained the task, the fewer possible solutions, hence the more convergent particular architectures (evolved or trained) to solve it become. The call it *a contravariance principle* (Cao & Yamins, forthcoming). I discussed briefly the assumption that biological systems are hard-pressed to perform efficient coding (Barlow 1978; Chirimuuta 2014). That is precisely because of the highly-constrained nature of their performance. If we accept it as a premise, we can see why we should in fact expect the artificial and biological nets to converge on similar solutions (provided we constrain them similarly).

That should raise our conviction in the fact that artificial nets do in fact provide good models for biological processes and that information-based explanations of perceptual processes would result in too many a system counting as representational. This perhaps also illuminates why the Moravec paradox happens to be the case. The intuitively simple perceptual task lay low only when we consider the ladder of abstraction. The conclusion that corresponds to the level of complexity is illusionary. Contravariance principle explains why these deceptively simple tasks are in practice extremely hard – precisely because they are in fact much more constrained.

There is a markedly strong support for the important informational similarity between biological and artificial nets (convergence) with the findings of ability of neural nets trained on visual classification tasks to be able also to predict neural (biological) activity with relatively high accuracy (Yamins & DiCarlo 2016). Moreover, Bashivan et al. (2019) developed an approach called „deep image synthesis" that is used to use machine learning systems to *generate* patterns of visual stimulation that maximally activate individual neurons in monkey's cortex (V4). The performance of the model was so high that it was able to generate stimuli that resulted in rates of neural firing in these neurons that were far greater than those evoked by any available natural image stimulus. (Poldrack, in press) A similar idea of identifying the interpretable building-blocks that maximally-activate a particular neuron to explain its function is also making a splash in 'explainable machine learning' (Olah et al. 2018).

What's equally important in our ability to localize and manipulate (e.g. through increasingly potent optogenetic or ablational techniques) neural representations is the idea that we seem to start being able to compare particular representational structures between different agents. Two especially impressive examples come from Schalk et al. (2017) and Marshel et al. (2019) study that seem to be able to optogenetically 'incept' a certain (carefully designed) representation with the brain. The former study was able to invoke face-like percepts in human subjects overlaid on inanimate objects. The latter study achieved successful substitution of visual stimuli that guided mouse behavior without the loss of ability to perform a high-order function.

We have analyzed evidence from conceptual analysis, similarity analysis, neuronal manipulation from both human, animal and artificial systems. We have seen that this data not discloses striking functional and structural similarities, enabling both comparison, but also mutual prediction. It points to the structural similarities of the different kinds of agents. We seem to be zoning in on resolving the dilemma offered by Ramsey in his discussion on the criteria of effective representation, whether 'are there mindless systems in which an internal element is performing a role that is most naturally (or intuitively, or justifiably, or beneficially) viewed as representational in nature?' (Ramsey 2007:196). Recent advances on the successful mutual comparisons and predictions from diverse (artificial and biological) systems lend support to there are deep structural similarities that explain the converging representational patterns as optimal coding procedures that effectively maximize 'output-to-scaffolding' ratio, very much compression-like. Perhaps that is plausible evidence for the existence of structural similarities not only between the environment and the agent but also between the agents themselves.

Structural similarity is an important feature for many theorists for explaining the representational capacities of the system (Cummins 1996). They successfully represent because there exists a systematic morphism (iso- or homomorphism) between important parts of the environmental makeup that are mirrored within the agential makeup that is in the business of faithfully representing its surroundings. This relation can be perhaps fruitfully described in the form of infomorphisms, as suggested by Barwise and Seligman in their 'Information Flow: Logic of Distributed Systems' (1997). Perhaps the notion of 'structural similarity' (Gładziejewski 2015) between the environment and the representational system and the idea of S-representation (Gładziejewski & Miłkowski 2017) can be mobilized for this purposes. This is a promising line of research, but the so far – the one for the future.

5. Conclusions

This paper was concerned with how artificial and biological systems share not only the feature of being information-processors, but also the challenge of performing this processing in a highly-selective, regimented way. This kind of outlook on the environment as structured, and as such predicable (and the informational reservoir it constitutes) confirmed by the findings of Natural Scene Statistics paradigm invited thinking of perception as compression-like. I argued that such 'clever' strategies are not only beneficial or optimizing, but necessary from both ecological (biological constraints) and computational (curse of dimensionality) standpoint. Any successful information-processing simply does not afford the luxury of total tracking. I also briefly reviewed the potential objections to informational-accounts as lacking content and worries about the unstable status of the notion of information itself (especially in neural contexts). It was conceded that although the objections of skeptics are well-motivated, they can be alleviated by existing theories of naturalization of informational content. It was argued that the postulated necessity of taking advantage of regularities in conjunction with content has to be interpreted representationally in the light of recent riveting breakthroughs in observability and manipulability of neural representations, their computational modeling in artificial neural nets. Finally, I tried to demonstrate the ability of mutual modeling and prediction between biological and artificial neural structures points to the possibility of establishing a structural similarity on the basis of the notion of infomorphism.

BIBLIOGRAPHY

Allen-Zhu, Z., Gelashvili, R., Micali, S., & Shavit, N. (2014). Sparse sign-consistent Johnson–Lindenstrauss matrices: Compression with neuroscience-based constraints. *Proceedings of the National Academy of Sciences*, *111*(47), 16872–16876.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183.

Baggio, G. (2018). *Meaning in the Brain*. MIT Press.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *1*(4), 371–394.

Barwise, J., & Seligman, J. (1997). *Information flow: The logic of distributed systems* (Vol. 44). Cambridge University Press.

Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis.

Science, *364*(6439), eaav9436.

Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological Review*.

Bechtel, W., & Shagrir, O. (2015). The Non-Redundant Contributions of Marr's Three Levels of Analysis for Explaining Information-Processing Mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322.

Bergstrom, C. T., & Rosvall, M. (2011). The transmission sense of information. *Biology & Philosophy*, *26*(2), 159–176.

Blinowska, K. J., Kuś, R., & Kamiński, M. (2004). Granger causality and information flow in multivariate processes. *Physical Review E*, *70*(5), 050902.

Block, N. (2014). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research*, *89*(3), 560–572.

Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nature Neuroscience*, *2*(11), 947–957.

Cantwell-Smith, B. (1996). *On the origin of objects*. MIT Press.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51.

Chaitin, G. J. (1977). Algorithmic information theory. *IBM Journal of Research and Development*, *21*(4), 350–359.

Chemero, A. (2009). *Radical embodied cognition*. Cambridge, MA: MIT Press.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese*, *191*(2), 127–153.

Clark, A. (1997). *Being there*. MIT Press Cambridge, MA.

Colgate, S. A., & Ziock, H. (2011). A definition of information, the arrow of information, and its relationship to life. *Complexity*, *16*(5), 54–62.

Collier, J. (1990). Intrinsic information. *Information, Language and Cognition: Vancouver Studies* *in Cognitive Science*, *1*, 390–409.

Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

Cummins, R., Putnam, H., & Block, N. (1996). *Representations, targets, and attitudes*. MIT press.

Dennett, D. C. (1991). Real patterns. *The Journal of Philosophy*, *88*(1), 27–51.

Dretske, F. (1982). The informational character of representations. *Behavioral and Brain Sciences*, *5*(3), 376–377.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Josa a*, *4*(12), 2379–2394.

Fodor, J. (1999, September 30). Jerry Fodor · Diary: Why the brain? · LRB 30 September 1999. *London Review of Books*, *21*(19). https://www.lrb.co.uk/the-paper/v21/n19/jerry-fodor/diary

Fodor, J. A. (1990). *A theory of content and other essays.* The MIT press.

Földiák, P., & Young, M. P. (1998). *Sparse coding in the primate cortex, The handbook of brain theory and neural networks*. MIT Press, Cambridge, MA.

Fresco, N. (2014). *Physical computation and cognitive science*. Springer.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, *104*(1–2), 137–160.

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.

Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, *193*(2), 559–582.

Gładziejewski, P., & Mi\lkowski, M. (2017). Structural representations: Causally relevant and different from detectors. *Biology & Philosophy*, *32*(3), 337–355.

Godfrey-Smith, P., & Sterelny, K. (2007). *Biological information*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Griffiths, P. E. (2001). Genetic information: A metaphor in search of a theory. *Philosophy of Science*, *68*(3), 394–412.

Grünwald, P. D., & Grunwald, A. (2007). *The minimum description length principle*. MIT press.

Hacking, I., & Hacking, J. (1983). *Representing and intervening: Introductory topics in the* *philosophy of natural science*. Cambridge University Press.

Hatfield, G. (2002). *Perception as unconscious inference*. Wiley Online Library.

Helmholtz, H. von. (1867). *Treatise on physiological optics, vol. III, trans. And ed. JPC Southall*. Dover.(Translated from the 3rd German edition, English edition 1962).[aJN, HER].

Huttegger, S. M., Skyrms, B., Smead, R., & Zollman, K. J. (2010). Evolutionary dynamics of Lewis signaling games: Signaling systems vs. partial pooling. *Synthese*, *172*(1), 177.

Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. Mit Press.

Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. MIT Press.

Hyvärinen, A., Hurri, J., & Hoyer, P. O. (2009). *Natural image statistics: A probabilistic approach to early computational vision.* (Vol. 39). Springer Science & Business Media.

Isaac, A. M. (2019). The semantics latent in Shannon information. *The British Journal for the Philosophy of Science*, *70*(1), 103–125.

Kanizsa, G. (1985). Seeing and thinking. *Acta Psychologica*, *59*(1), 23–33.

Kiefer, A. B. (2017). *Literal perceptual inference*.

Korbak, T. (2015). Scaffolded minds and the evolution of content in signaling pathways. *Studies in Logic, Grammar and Rhetoric*, *41*(1), 89–103.

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Le, Q. V., Karpenko, A., Ngiam, J., & Ng, A. Y. (2011). ICA with reconstruction cost for efficient overcomplete feature learning. *Advances in Neural Information Processing Systems*, 1017–1025.

Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). *JOSA A*, *22*(10), 2013–2033.

Lewis, D. (2008). *Convention: A philosophical study*. John Wiley & Sons.

Li, X., Lu, H., Zhang, L., Ruan, X., & Yang, M.-H. (2013). Saliency detection via dense and sparse reconstruction. *Proceedings of the IEEE International Conference on Computer Vision*, 2976–2983.

Lupyan, G. (2015). Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems. *REVIEW OF PHILOSOPHY AND PSYCHOLOGY*, *6*(4), 547–569.

MacKay, D. J., & Mac Kay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information, henry Holt and co. *Inc., New York, NY*, *2*(4.2).

Matsumoto, M., & Komatsu, H. (2005). Neural responses in the macaque V1 to bar stimuli with various lengths presented on the blind spot. *Journal of Physiology*, *93*(5), 2374–2387.

Mavlankar, A., Chang, C.-L., & Giros, B. (2006). Intraframe Prediction with Intraframe Update Step for Motion-Compensated Lifted Wavelet Video Coding. *Picture Coding Symposium*.

Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. MIT press.

Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., & Mordvintsev, A. (2018). The building blocks of interpretability. *Distill*, *3*(3), e10.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607.

Orlandi, N. (2014). *The innocent eye: Why vision is not a cognitive process*. Philosophy of Mind.

Orlandi, N. (2016). Bayesian perception is ecological perception. *Philosophical Topics*, *44*(2), 327–352.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT press.

Piccinini, G. (2015). *Physical computation: A mechanistic account*. OUP Oxford.

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(3), 341–365.

Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge University Press.

Rathkopf, C. (2017a). Neural information and the problem of objectivity. *Biology & Philosophy*, *32*(3), 321–336.

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

Schalk, G., Kapeller, C., Guger, C., Ogawa, H., Hiroshima, S., Lafer-Sousa, R., Saygin, Z. M., Kamada, K., & Kanwisher, N. (2017). Facephenes and rainbows: Causal evidence for functional and anatomical specificity of face and color processing in the human brain. *Proceedings of the National Academy of Sciences*, *114*(46), 12285–12290.

Sekuler, A. B., & Palmer, S. E. (1992). Perception of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: General*, *121*(1), 95.

Shannon, C. E. (1956). The bandwagon. *IRE Transactions on Information Theory*, *2*(1), 3.

Shea, N., Godfrey-Smith, P., & Cao, R. (2017). Content in simple signaling systems. *The British Journal for the Philosophy of Science*.

Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford University Press.

Thomson, E., & Piccinini, G. (2018). Neural representations observed. *Minds and Machines*, *28*(1), 191–235.

Timme, N. M., & Lapish, C. (2018). A tutorial for information theory in neuroscience. *ENeuro*, *5*(3).

Tishby, N., & Zaslavsky, N. (2015). Deep learning and the information bottleneck principle. *Information Theory Workshop (ITW), 2015 IEEE*, 1–5.

Tootell, R. B., Switkes, E., Silverman, M. S., & Hamilton, S. L. (1988). Functional anatomy of macaque striate cortex. II. Retinotopic organization. *Journal of Neuroscience*, *8*(5), 1531–1568.

Wilkenfeld, D. A. (2019). Understanding as compression. *Philosophical Studies*, *176*(10), 2807–2831.

Wolff, J. G. (1982). *Language acquisition, data compression and generalization*. Pergamon.

Wolff, J. G. (1993). Computing, cognition and information compression. *AI Communications*, *6*(2), 107–127.

Xu, L., Zheng, S., & Jia, J. (2013). Unnatural l0 sparse representation for natural image deblurring. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1107–1114.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356.