Review of Philosophy and Psychology (2020) 11:507–515
https://doi.org/10.1007/s13164-020-00500-x

Editorial: Computationalism Meets the Philosophy of Information

Brian Ball[1,2] · Fintan Nagle[3] · Ioannis Votsis[1]

Published online: 14 August 2020

1 Background: Minds, Brains and Computers

That the human mind, or at least the brain, is a kind of machine is a view that goes back a long way. Descartes, for instance, thought that bodies including brains are functionally no different than clocks and explained memory and other mental processes in purely mechanical terms. The most sophisticated machines ever built by humans are computers. They are also the machines that are of most benefit to us, helping us solve problems all the while exhibiting behaviour that begs to be interpreted as intelligent. It should thus come as no surprise that views of the mind, or at least the brain, have been increasingly conceived of in computational terms.

McCulloch and Pitts (1943) suggested in an early influential paper that the mind may be something like a Turing machine. In other words, that a mind can take the form of a machine or an algorithm, a view famously defended by Turing himself (Turing 1950). This view came to be known as 'classical computationalism'. It was quickly met with an onslaught of objections, and in reaction a number of liberalisations ensued – see, for example, Scheutz (2002).

One version that has recently been gaining ground attempts to articulate the notion of computation in terms of the notions of information and information-processing. The latter two have received formal treatments in the field of information theory, pioneered by Shannon (1948) with his mathematical theory of communication, which asked how to send messages efficiently over a noisy channel. What is now called 'Shannon information' is a probabilistic measure of the uncertainty associated with the value of a given message. The measure is essentially Gibbs' entropy formula, transposed from the domain of statistical thermodynamics to that of information theory. Philosophers have not only taken note of these treatments but have also sought to develop a novel

✉ Ioannis Votsis
ioannis.votsis@nchlondon.ac.uk

[1] New College of the Humanities, London, UK

[2] University of Oxford, Oxford, UK

[3] Imperial College, London, UK

branch of philosophy to make use of them. The emergence of the philosophy of information has seen a flurry of publications which seek to relate it to traditional philosophical problems – see, for example, Floridi (2016).

Interest in these two areas, i.e. computationalism and the philosophy of information, is on the ascendancy. This special issue is devoted to the intersection between them, especially to papers that engage in a meaningful way with recent work in cognitive science. Four invited papers (of which three are individually authored, namely by Rosa Cao, Michael Rescorla and Mark Sprevak, and one is co-authored, namely by Nir Fresco, Simona Ginsburg and Eva Jablonka) as well as four contributed papers (of which three are individually authored, namely by Francis Fallon, Corey Maley and Stephen Mann, and one is co-authored, namely by Marc Artiga and Miguel Angel Sebastian) make up the collection. In what follows, we offer short summaries of these papers, first the invited ones and then the contributed ones but always proceeding in an alphabetical manner. We then present some brief remarks on what we think has been learnt since the early days of computationalism.

2 Contents of the Special Issue

In her contribution, **Rosa Cao** casts doubts on the novelty of predictive models of perception (particularly those purporting to give a semantic story of what goes on in visual information-processing). The advocates of such models take perceptual information-processing to be an active and internally-driven process. According to this account, the brain does not merely react to external stimuli but rather produces top-down predictions (about the possible causes of the stimuli) which get compared with the bottom-up sensory signals. The only bottom-up signals left travelling to the higher areas of the brain are the prediction errors, i.e. the discrepancy between the predictions and the sensory signals, which our brain continually seeks to minimise. The punchline here is that perceptual information content, understood in terms of notions such as Shannon information, is internally, not externally, driven.

Cao argues that predictive models only offer an 'interpretative gloss' on the information-processing of neural signals in perception. Traditional models, she claims, are not inextricably tied to a bottom-up and passive view of perceptual information-processing, as advocates of predictive models suggest. Rather, on more sophisticated accounts of traditional models, external stimuli may initiate a cascade of neural activity from lower to higher areas of the brain but they can also be modulated by various top-down processes like adaptation, attention and memory. Indeed, Cao argues, the same information can be gleaned under both families of models and, as such, the families are informationally equivalent. It's just that what the predictive model theorists call 'the error signal', the traditional model theorists call 'the bottom-up input signal'.

Moreover, no differences in the anatomy of the brain can be justifiably attributed to one family of models but not the other. Indeed, any differences found in specific models of either family are due to auxiliary assumptions about mappings that each makes, but nothing prevents these assumptions (mutatis mutandis) from accompanying models of the other family. So, Cao reasons, any pretence that the two families of models make different predictions about information flow in perception is easily squashed. The two families of models, she even goes as far as suggesting, can be thought of as not only empirically but also theoretically equivalent.

In their contribution to this special issue, **Nir Fresco, Simona Ginsburg, and Eva Jablonka** offer a taxonomy of functional information (ToFI) for use in e.g. cognitive science and the theory of animal communication. The notion of information they employ is neo-Peircian: 'information is a triadic relation amongst a receiver, a difference-maker…, and an object/feature/event/state of affairs.' It is also functional: information is something that makes 'a *systematic, causal* difference to the… [receiver's] goal-directed behaviour'. One intriguing upshot of this receiver-relative, functional characterization of information is (the authors claim) that the notion of a 'vehicle' that carries it must be jettisoned, since this presupposes (falsely on this view) that information is a 'commodity' that is 'mind-independent' and can be 'carried' or 'conveyed'.

In the animal communication literature, a cue is (roughly speaking) a feature of the world that can be used by an agent or receiver to guide actions; and a signal is a cue that has evolved (in a sender) to be used (by a receiver) in this way. ToFI extends the cue-signal distinction (at least approximately) by drawing a further distinction on each side of this initial divide. Its four central notions are: *datum* (a kind of '*potential* information' for a given receiver); *sign* (a refinement of the notion of a cue, and a kind of *actual* information); *signal* (a sign produced by a sender that evolved to produce it); and *symbol* ('an intentional signal that is part of a systematic, rule-governed, self-referential-signalling system'). Fresco et al. are at pains to stress that the evolutionary processes they appeal to need not be phylogenetic but can be ontogenetic (as in learning) or cultural (as in the development of a convention).

The approach taken by these authors also allows them to resolve a certain problem surrounding the distinction between natural and non-natural information. Roughly, the worry is that these do not appear to be exclusive and exhaustive categories: vervet alarm calls, for instance, do not function on a merely correlational basis (as is required for so-called 'natural; information); yet they are not intentional (in the manner required for Gricean non-natural meaning). The triadic, functional approach to information embodied in ToFI recognizes that there are both 'natural' and 'non-natural' components here: roughly, the first consists in the correlation between the sign (or in this case signal) and the situation (to which it functionally refers); while the second involves the functional change it brings about in the receiver (a kind of interpretive inference). There is plenty more in this highly textured paper: for instance, the authors devote a section to comparisons with the views of Skyrms and Corning, and they respond to a pair of objections in another; they also relate their work to that undertaken by others (including Dretske and Scarantino).

Michael Rescorla focuses his contribution on the computations that perceptual systems perform over mental representations, particularly the kinds of computations involved when combining proximal sensory input or cues to estimate distal conditions or properties. The main thesis of the essay is that one and the same distal condition, e.g. size, may be represented by distinct (inter-modal or intra-modal) perceptual states. Such representations are thus, in Rescorla's own words, 'co-referring'. Rescorla claims that support for this view comes from Bayesian sensory cue combination models of the perceptual system, particularly coupling prior models. He takes such models to offer our best explanation of the said phenomena. The perceptual system is thought of as encoding prior probabilities and likelihoods which in turn allow it to compute the posterior probabilities that a specific distal condition or property holds given a certain sensory input or cue. Several such computations occur until a privileged hypothesis - which he takes to be a perceptual representation - is chosen on the basis of posterior probability maximisation.

Among other things, Rescorla connects his views with two key figures from the history of philosophy, Berkeley and Frege, though in both cases some glossing over is essential. Berkeley, like Rescorla, seems to argue that different modalities yield distinct perceptual ideas; the gloss here being that perceptual ideas are identified as perceptual representations. Frege, like Rescorla, seems to argue that one may present the same things in different ways or modes; the gloss here being that distinct modes of presentation are identified as distinct perceptual representations. Rescorla takes this Fregean connection to provide an important moral in providing explanations in perceptual psychology: such explanations must, at least sometimes, take into account distinct perceptual representations as these influence the inferences drawn by the perceptual system. The essay concludes with a defence of the view that such perceptual representations are relatively coarse-grained so that each one can persist even in the light of changing priors and considerable variation in sensory input or cues.

Mark Sprevak, in his contribution, looks at the relationship between information and representation in models of cognition. In particular, he compares traditional information-theoretic models with more recent models. His main thesis is that endorsement of the more recent models inevitably brings with it an acceptance of not only the traditional account of Shannon information involvement in cognition but also a conceptually novel and logically distinct account of this involvement. The traditional account, it is claimed, feeds off of the capacity of states as representational vehicles.

To support this claim, Sprevak reviews a number of influential information-theoretic approaches to semantics and representation and concludes that Shannon information measures emerge in relation to neural (and environmental) states simply because such states, qua representational vehicles, possess the right probabilistic properties. To be exact, they allow the construction of probabilistic ensembles which in turn allow the construction of associated measures of Shannon information. The second account of Shannon information involvement in cognition, it is claimed, utilises the representational content of neural states.

To support this claim, Sprevak considers recent developments in models of cognition, e.g. Bayesian models. Unlike in the case of traditional models, the probability distributions here are not associated with a neural state occurring but rather with the brain’s estimates of certain such states occurring. Like in the case of traditional models, the right probabilistic properties are present and therefore give rise to Shannon information measures. Sprevak clarifies that there is no reason why a model of cognition could not incorporate both the traditional and the newer account of Shannon information involvement in cognition. Moreover, he argues that the reason why the two accounts are conceptually and logically distinct is because they typically concern different sets of outcomes, probability values and even kinds of probabilities. Sprevak concludes his essay by offering some fruitful suggestions of how the two accounts may potentially be connected.

As **Marc Artiga and Miguel Angel Sebastian** point out in their paper, a theory of (the nature of) representation must address two questions, namely: (the semantic question) What determines the content of a given representation? and (the metasemantic question) What makes a given state a representation in the first place?

One strategy for providing a naturalistic theory of representation has been to appeal to the (causal-cum-correlational) notion of information. In particular, a number of recent Scientifically Guided Informational Theories (SGITs) - specifically, those advocated by Eliasmith and Usher - hold that a given state represents a given content iff: (1) its probability of obtaining is higher given that the content state obtains than that any other alternative content state does (the forwards condition); and (2) the probability of the content states obtaining is higher given that it obtains than given that any other alternative representation does (the backwards condition). And Rupert requires the first condition alone, provided that the represented contents constitute natural kinds.

Artiga and Sebastian argue, however, that these theories face four specific problems, and moreover, that these problems are not superficial, but instead suggest that some non-informational notion must be appealed to in explaining representation.

The first problem is the problem of error. Dretske, who advocated an early informational theory of representation, faced an acute version of this problem, since he required the probability of the content state to be equal to one, given the presence of the representation, thereby precluding misrepresentation. But while SGITs overcome this version of the problem by relaxing Dretske's condition, and requiring only that representations are the states that are most highly correlated with their contents, Artiga and Sebastian argue that they nonetheless 'cannot account for some cases in which misrepresentation is more frequent that accurate representation'.

The second ('distality') problem is that SGITs are likely to predict that the representational content of a given state is too proximal – for instance, certain states involved in late visual processing may be taken to represent states in early visual processing with which they are very highly correlated, or 'face-looking' things, rather than faces. Third, SGITs can accommodate neither 'ambiguous' representations with e.g. disjunctive contents, nor representational redundancy (with a single state represented more than once e.g. in different sensory modalities). Finally, SGITs fail to even address the metasemantic question, which is nonetheless pressing. By way of diagnosis, Artiga and Sebastian repeatedly suggest that an appeal to the notion of function might help to supplement - or possibly even replace - the notion of information in an adequate theory of representation.

Francis Fallon's piece is concerned with consciousness, and in particular with Searle's view, the Integrated Information Theory (IIT) of Tononi and Koch, and the relations between them. While he disagrees with both theories, Fallon thinks they have more in common than Searle recognizes, and that they pose a challenge which even their opponents must meet.

Fallon begins by introducing IIT. According to this theory, conscious systems *exist* (and so have 'cause-effect power') *intrinsically* (i.e. from their own perspective); moreover, they *specify information*; and they do so via *integration* (in a unitary way). Any system meeting these four conditions (which requires a reentrant architecture, comprising feedback loops in which output serves as input) counts as integrating information; but this does not yet suffice for consciousness. IIT provides a measure (phi) of the quantity of integrated information, and identifies a structure's being conscious with its having 'the local maximum… of the system's integrated information' (because consciousness is definite).

Searle engages with IIT, but according to Fallon he 'misunderstands the central notion of integrated information', conflating it with Shannon information. As Fallon sees it, IIT and Searle are alike in regarding the latter as merely extrinsic, or observer-relative. Moreover, IIT locates intrinsic information in the possession of appropriate causal powers, just as Searle suggests: what's crucial, according to IIT, is the causal dynamics of reentrant systems – a feature that is substrate neutral, in line with Searle's official view; and the theory provides a response to the charge of mysticism that Searle has faced.

Next, Fallon explores the ontologies of the two theories. While Searle recognizes both intrinsic existence (e.g. in the case of fundamental physical particles) and intrinsic intentionality (which is intimately bound up with consciousness in his view), IIT affords ontological priority to conscious systems, which alone have intrinsic existence. Searle's ontology allows him to raise an arbitrariness concern against certain rival theories: '[a]ny object under the right description can be regarded as a digital computer implementing a program' that processes information; implementing a program and processing (Shannon) information are (extrinsic) observer-relative notions, and the observer can choose an arbitrary description. But despite its different ontology, 'IIT joins Searle in challenging information processing and computational theories of mind on the grounds of arbitrariness.' Since it grounds existence in causal integration, whatever lacks such integration is (intrinsically) no system at all, hence a fortiori not an information processing computational system. Thus, both theories, in their own ways, raise 'the arbitrariness question: What makes for a system in the first place?' This is a question, on Fallon's view, that requires a response, even from the opponents of Searle and IIT.

Corey Maley looks at a key application of information theory to biological systems: neural networks in the brain. After giving an accessible introduction to Shannon information and entropy, Maley observes that the behaviour of a neuron cannot be summed up as a list of spikes (action potentials) occurring at particular times. Naive analyses record an identical data point each time a neuron fires. However, real neurons do not always spike identically: their peak voltages and decay rates can vary, which is easily visible in the shape of the spike on a continuous recording. This means that errors (missing spikes or recording false positives) are inevitable when converting a raw recording into a list of spikes.

Maley argues that neural activity is a key target for analysis by information theory on continuous signals. This view sees the brain more as a collection of oscillating systems than as a network of spiking neurons. It comes to the fore as we discover more about the long-range communication and synchronisation of groups of oscillating neurons - the rhythmic firing that generates EEG signals. On a smaller scale, it is now clear that within a neuron, individual dendrites also produce tiny spikes, performing computations which are even more local than the action potential.

Stephen Mann discusses the philosophical interpretation of Shannon's communication theory and the statistical machinery surrounding the vexing concept of information. In Mann's view, Dretske failed to appropriately generalise communication theory, leading to an inadequate contemporary causal interpretation of information. Instead, Mann argues for an alternative functional interpretation which engages better with experimental results.

Mann's interpretation deals with the multiple representations of a message, which is at once an electrical (or biological) signal, a string of encoded symbols or signs, and a carrier of meaning for its receiver: the "user" of the information it contains. Shannon's theory deals only with real-world signals and their symbol strings, aiming to "[reproduce] at one point either exactly or approximately a message selected at another point". Mann is not content with the commonly-accepted divide between "Shannon information" (codes) and "semantic information" (that which is expressed by codes). Mann shows that Dretske's interpretation of communication theory was driven by the need to set objective probabilities in order to provide epistemological grounding. This provides what Mann calls "a user-independent resource at the foundation of naturalist epistemology."

This setup is vulnerable to the reference class problem: one cannot work out the probability of a token without knowing (and being able to count) the class to which that token belongs. The functional interpretation addresses the reference class problem by defining information as relative to a user's background knowledge. What an agent already 'believes' will determine the quantity of information it can receive from a given signal. In order for an agent to extract information from a signal, it must be configured to expect that information: to a trained rat, the ring of a bell feels very different compared to an untrained rat. Mann argues that cognitive science and microbiology are applying these ideas already, and that philosophy of information needs to catch up.

Mann draws parallels between functional information, decision theory, and Bayesian inference. His interpretation challenges the usual distinction between causal and semantic information, working against the irrelevance claim: the idea that statistical information has no bearing on the meaning a signal has for the receiver. Developing his reading in the context of cellular signalling and perceptual decision tasks, Mann aims to integrate communication theory better into the philosophical toolkit.

3 What Have we Learnt since Turing (1950)?

Though it is near impossible to summarise all the pertinent developments since Turing's seminal paper, we will at least attempt to touch upon some key issues. Besides the emergence of Shannon information (in fact, two years earlier), the intervening years have seen the introduction of various other influential notions and distinctions. Here we restrict our comments to two such notions of information. The first is that of semantic information. It arguably originates in the work of Bar-Hillel and Carnap (1952) who argued that information and truth are intimately tied together. More specifically, they supported the claim that the informational content of a proposition is inversely proportional to the likelihood of its truth. In recent years, fruitful connections have been made between this account and our understanding of truth via possible world semantics. For example, informative propositions are said to help narrow down the actual world from the set of possible worlds (see, for example, Stalnaker 1984 and Sequoiah-Grayson 2007).

Yet another notion worth mentioning is that of Kolmogorov complexity. This notion sits at the heart of algorithmic information theory. The idea behind it is quite simple.

We measure the information that a string carries in terms of the shortest computer program (in some designated universal Turing machine) that yields that string. The notion gets its name from the work of Andrey Kolmogorov (1965) but at least one other figure deserves credit for it. Ray Solomonoff, a one-time student of Carnap, proposed that simpler hypotheses are more likely to accurately predict the future than more complex ones and should therefore be assigned higher prior probabilities, where hypotheses can be thought of as sequences of symbols that extend the existing evidence we have which are also sequences of symbols. Once again, fruitful connections have been made between such views and various philosophical subjects, most notably confirmation theory and the philosophy of statistics (see, for example, Bandyopadhyay and Forster 2011 and Votsis 2016).

All three notions, i.e. Shannon information, semantic information and algorithmic information, have proved fertile grounds upon which to sow various useful accounts of cognate notions like truth, knowledge and belief. The trouble is that the links between the three notions themselves are not always fully understood. At least some of the work presented in this special issue has sought to bridge this gap in our understanding. As we saw in the contributions of Cao, Mann, Rescorla and Sprevak, the similarities and differences between semantic and probabilistic conceptions of information remain underexplored and are ripe for further treatment. The same can be said for the similarities and differences between the semantic and algorithmic conceptions of information. At the same time, some theorists – including Fresco, Ginsburg, and Jablonka, as well as Artiga and Sebastian, in this special issue – have looked to pursue a pragmatic conception of information. Following Bateson's (1972) REF idea that a bit of information is a 'difference that makes a difference', the key thought here is that the use to which an agent is able to put a correlation is crucial in determining its status as information.

Beyond our cursory discussion of the three notions of information, there remains the question of the extent to which the research carried out since Turing's seminal work has shed any light on how minds are embodied. Progress has been fraught with setbacks. As Maley notes in his contribution, even individual neurons are extremely complex entities and pose a challenge to decipher. Indeed, notions that encode higher levels of information – for example, cognition, self-perception, affect, and notoriously, consciousness – have resisted in-depth treatments. Kindred notions, e.g. those employed in accounts (such as IIT) that seek to explain phenomena like consciousness, can be subjective and difficult to communicate, as Fallon argues with Searle's apparent misunderstanding of the concept of integrated information.

Yet we cannot ignore that positive steps in our understanding of the brain's mechanics have also been taken. Neuroscience has met with enormous success in proposing computational elements that may underlie perceptual processing, such as the perception of motion, frequency, or rhythm. Moreover, as Artiga and Angel Sebastian suggest, many questions around representation, information, and embodiment are becoming clearer. The contributions of Mann and Maley both show that information theory is already a useful tool for systems biology and the brain sciences. Current research programmes on embodiment (Buzsáki 2006, EEG) and statistical surprise (Friston 2010, fMRI) continue to drive the integration of concepts such as free energy into philosophy. Turing had a thorough biological grounding and worked near the end of his life on morphogenesis. He would have been fascinated by our understanding of the molecular, cellular and neural circuitry that underlies the mind – but perhaps disappointed at our relative lack of progress in how precisely the mind is embodied.

References

Bandyopadhyay, P., and M. Forster, eds. 2011. *Handbook for the Philosophy of Science*. Vol. 7: Philosophy of Statistics. Amsterdam: Elsevier.

Bar-Hillel, Y., and R. Carnap. 1952. *An outline of a theory of semantic information. Technical Report No. 247, Research Laboratory of Electronics*. Cambridge: MIT.

Bateson, G. 1972. *Steps to an ecology of mind*. San Francisco: Chandler Publishing Company.

Buzsaki, G. 2006. *Rhythms of the brain*. New York: Oxford University Press.

Floridi, L., ed. 2016. *The Routledge handbook of philosophy of information*. London: Routledge.

Friston, K. 2010. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience* 11 (2): 127–138.

Kolmogorov, A.N. 1965. Three approaches to the quantitative definition of information. *Problems of Information Transmission* 1 (1): 1–7.

McCulloch, W., and W. Pitts. 1943. A logical Calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 7: 115–133.

Scheutz, M. 2002. *Computationalism: New directions*. Cambridge: MIT.

Sequoiah-Grayson, S. 2007. The Metaphilosophy of information. *Minds and Machines* 17: 331–344.

Shannon, C.E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27 (3): 379–423.

Stalnaker, R. 1984. *Inquiry*. Cambridge, MA: MIT Press.

Turing, A.M. 1950. Computing machinery and intelligence-AM Turing. *Mind* 59 (236): 433–460.

Votsis, I. 2016. Philosophy of science and information. In *The Routledge handbook of philosophy of information*, ed. L. Floridi. London: Routledge.