Chapter 2
Computing Mechanisms and Autopoietic Systems

Joe Dewhurst

Abstract This chapter draws an analogy between computing mechanisms and autopoietic systems, focusing on the non-representational status of both kinds of system (computational and autopoietic). It will be argued that the role played by input and output components in a computing mechanism closely resembles the relationship between an autopoietic system and its environment, and in this sense differs from the classical understanding of inputs and outputs. The analogy helps to make sense of why we should think of computing mechanisms as non-representational, and might also facilitate reconciliation between computational and autopoietic/enactive approaches to the study of cognition.

2.1 Introduction

Computational and autopoietic (or enactive[1]) approaches to cognition have traditionally been opposed to one another, primarily due to a disagreement about whether or not representation is necessary for cognition. On the one hand, computation has classically been understood as inherently representational (Sprevak 2010: 260), leading to the conclusion that if cognition is computational then it must also be representational (Fodor 1981: 180). On the other hand, autopoietic theory and the enactive approach that it inspired have both argued that cognition does not require representation (see e.g. Varela et al. 1991: 8; Thompson 2007: 52–3). This has led to a situation in which we have two divergent approaches to the study of cognition that appear to be fundamentally irreconcilable.

[1]Here I have in mind the "biological enactivism" of Varela (e.g. 1991), Thompson (e.g. 2007), and di Paolo (e.g. 2005), as opposed to the "sensori-motor" enactivism of Hurley (e.g. 1998), Noë (e.g. 2004), and Hutto and Myin (2013). See Villalobos and Ward (2014, fn 1) for a discussion of this distinction. All further references to enactivism should be understood as referring to biological enactivism.

J. Dewhurst (✉)
Department of Philosophy, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK
e-mail: joseph.e.dewhurst@gmail.com

V.C. Müller (ed.), *Computing and Philosophy*, Synthese Library 375,
DOI 10.1007/978-3-319-23291-1_2

Recently, however, there have been several attempts to give non-representational accounts of computation (see e.g. Egan 1995; Piccinini 2008), which could in theory lead to reconciliation with autopoietic/enactive approaches. One such attempt is Gualtiero Piccinini's mechanistic account of computation, which characterises computational states as components in a mechanism (see his 2007). According to this account it is not necessary that computational states represent, although it does not rule out this possibility either – rather the question of what it means to compute becomes separated from that of what it means to represent (Piccinini 2004: 404). Similarly, whilst an autopoietic system might be interpreted as representing, this is neither essential to its identity nor constitutive of its operation (Maturana and Varela 1980: 78). Here I draw an analogy between the two kinds of system, based on this point of similarity. My primary aim is to elucidate the nature of inputs and outputs in the mechanistic account, but I also hope that this comparison might facilitate reconciliation between computational and autopoietic/enactive approaches to the study of cognition. Given the current dominance of computationalism in cognitive neuroscience, this would allow for autopoietic and enactive approaches to make a more meaningful contribution to practical research.

The first two sections will introduce computing mechanisms and autopoietic systems, respectively. The third and fourth sections will expand on the analogy between the two, focusing first on the role of representations and then turning to inputs, outputs, and perturbations. Finally there will be a brief discussion of how this analogy might help reconcile the two approaches, and how this might be of benefit to the study of cognition.

2.2 Computing Mechanisms

Classical accounts of computation have tended to invoke representation in order to distinguish computational states from mere physical states (Ramsey 2007: 43), and also to individuate those states (Sprevak 2010: 24–6). This is problematic if you have prior theoretical objections to representation; such as thinking that representation introduces a vicious circularity into computational explanation (Piccinini 2004: 377). It is concerns of this kind that have prompted Piccinini to develop a non-representational account of computation (see his 2008), although it should be noted that Piccinini is not committed to a totally non-representational account of cognition outwith its computational elements.

Piccinini's account is inspired by recent work on mechanistic explanation in cognitive science (2007: 501). Mechanistic explanation focuses on describing a target phenomenon in terms of the structured interaction of physical components (Craver and Bechtel 2006), where these components are understood as fulfilling certain functional roles. It is claimed that mechanistic explanation is especially suited to the special sciences, such as cognitive science and computer science,

where the traditional deductive-nomological model might be less relevant, as we are unlikely to discover broadly applicable natural laws (Bechtel 2005: 208; Craver and Darden 2005: 234).

A computing mechanism[2] is defined as a physical system that carries out systematic transformations on strings of digits (this does not rule out its also having some other function, such as controlling a system as a result of these transformations). By 'systematic transformation' I simply mean any physical interaction that will transform strings of digits in a way that would be replicated if the same kind of interaction were to take place under relevantly similar circumstances. This is distinct from the view that computation is simply an implementation of a systematic transformation from input to output. In contrast, the mechanistic account defines computation as the systematic transformation of digits, and does not in fact require an input or output of any kind (see below).

This structure requires a minimum of two components: a string of digits and a processor to transform those digits. It may also include an input device (for transforming external stimuli into strings of digits), an output device (for reversing this process), and a memory component (which may just be a looped string of digits). Whilst many computers will include these additional components, the simplest forms of computation can proceed without them (Piccinini 2007: 514). Each component is individuated functionally, such that every digit of the same type is treated in the same way by the mechanism, and every processor of the same type performs the same transformation on any given string (Piccinini 2007: 508–20). Whilst this processing of digits can be given a representational interpretation, it is not necessary to do so in order to explain how the mechanism functions (see Piccinini 2008).

This can be contrasted with what Sprevak calls "the received view", which is that computation must necessarily invoke representation in order to individuate digits and processors (2010: 260). Typically this is cashed out in one of two ways: representation can be required either for the individuation of computational states and processes (this is Sprevak's preferred interpretation), or it can be required in order to explain the transformation from input to output. In both cases the vehicles are understood as being physical states of some kind (such as holes in a punch-card or variations in voltage level), whilst the content can be either features of the world or abstract entities such as numbers (again, Sprevak prefers the latter, more minimalist interpretation). Under the mechanistic account there is no need for any appeal to representational content for either individuation or causal explanation.

I will not provide any further defense of the mechanistic account at this time, although I hope that the analogy with autopoietic systems will help elucidate precisely why computing mechanisms are not intrinsically representational. For a more detailed defense of the mechanistic account see Dewhurst (2014).

[2]Strictly speaking, a digital computing mechanism, although similar accounts can be given for analog or generic computing mechanisms (see Piccinini and Bahar 2013).

2.3 Autopoietic Systems

Maturana & Varela's theory of autopoiesis was developed in response to a perceived need for a better definition of living systems, which they take to encompass cognitive systems (i.e. all cognitive systems will be living systems, even if not all living systems are cognitive systems). It does this by focusing on the homeostatic nature of living organisms, which seem to be uniquely capable of preserving their structural integrity in response to environmental interference (Maturana and Varela 1980: 78ff). This criterion seems to capture everything that we might instinctively think of as living, and has the added benefit of not automatically excluding non-organic (or even non-physical) systems.

The neologism *autopoiesis*, formed from the Greek *auto* ("self") and *poiesis* ("production") is intended to capture the essence of this criterion. An autopoietic system is one that is focused towards continual self-production, as opposed to an allopoietic system, which produces something other than itself (Maturana and Varela 1980: 80–1). As a natural, deterministic entity an autopoietic system is non-teleological and therefore non-representational. It also exhibits conceptual and physical unity (*ibid*: 96): conceptual unity because self-production grants it an identity independent of any external observer, and physical unity because of its ability to maintain a coherent structure over time (*ibid*: 97). This reinforces its non-teleological status, as its behaviour emerges out of homeostatic regulation rather than being directed towards any external goal. Maturana & Varela acknowledge that we will often describe an autopoietic system in teleological terms, but descriptions of this kind are to be understood as strictly observer-relative, rather than reflecting anything intrinsic to the system itself.[3]

The archetypal example of a physical autopoietic system is a living cell, which whilst being "a thermodynamically open system" is nonetheless a closed unity in the sense that it "produces its own components [...] in an ongoing circular process" (Thompson 2007: 98). The body, as a collection of autopoietic cells, is granted second-order autopoietic status, and cognition is seen as a function of the living organism as a whole (Maturana and Varela 1980: 13). One could even imagine autopoietic systems made up of collections of organisms, although Maturana & Varela disagreed about whether this would be possible, and so did not comment on it (1980: 118). Cognition, for Maturana & Varela, is not intrinsic to an autopoietic system, but is rather an observer-relative categorization of the kind of behaviour that it exhibits.

This is by necessity a very brief outline of Maturana & Varela's theory of autopoiesis, but it should be sufficient for current purposes. For an accessible

[3]Varela later turned away from the idea that autopoietic systems are non-teleological (see Weber and Varela 2002), but this position is maintained in Maturana's later work. I focus here on the formulation of autopoiesis given in Maturana and Varela (1980).

overview see Mingers (1989); for the original formulation see Maturana and Varela (1980). In the following sections I will provide additional clarification when it is required in order to make sense of the analogy with computing mechanisms.

2.4 Representation

Representations are commonly seen as an essential component of cognitive scientific explanation (at least on the classical view), and thus naturalising representation is seen as an essential task for the philosophy of cognitive science (cf. Ramsey 2007). This makes it all the more interesting that both the mechanistic account and autopoietic theory are claimed to be non-representational (Piccinini 2008; Maturana and Varela 1980: 91), whilst at the same time serving as the basis for theories of cognition.

Computing mechanisms are non-representational because their components (digits and processors) can be individuated without specifying representational content. This is done functionally, by describing how a given digit-type will behave when it encounters a given processor-type. For instance, imagine a system with digits of two types (call them 0 and 1), and a certain processor (call it an x-gate). This processor takes pairs of digits and produces a single digit, based on what the pair consists of. If the pair is 0–0, 0–1, or 1–0, it produces a 0, and if the pair is 1–1, it produces a 1. Note that whilst this appears to correspond precisely with the logical function AND, we do not need to know this in order to individuate the digits, meaning that it is unnecessary for us to say that it represents this function. Additionally, our choice to label the first kind of digit 0 and the second 1 was completely arbitrary, and we could just as easily have reversed it, in which case our processor would appear to correspond with the OR function. The physical process is not sufficient to determine what logical function is taking place. All that is required for the computing mechanism to operate correctly is a physical difference between the two kinds of digit that are recognised by the processor (this could be done in a variety of ways, such as voltage levels or the presence/absence of a hole on a tape). Whilst we might choose to attribute representational content to computational states or processes, this attribution is not what makes the system computational, nor is it essential to our understanding of what it is to compute (see Piccinini 2008 for more detail).

Autopoietic systems are non-representational because, as Maturana & Varela put it, “there is no specification in the cell of what it is not” (1980: 91). This is a consequence of what Maturana calls “structural determinism”, which is true of any physical system: the result of any interaction with a physical system is fully determined by the intrinsic structure of that system, meaning that anything external to the system is merely a trigger, rather than a determiner, of that system’s structural dynamics (see Maturana 2003: 61).

Representation is further ruled out by the lack of teleology described earlier – to represent requires being able to misrepresent, which implies some purpose or

intention by which to judge failure or success (Millikan 1995: 186). Whilst it "may be metaphorically useful" to attribute representational content to an autopoietic system, this is ultimately "inadequate and misleading" (Maturana and Varela 1980: 99). Representation, just like teleology, should be treated as an epistemic tool for the benefit of an observer, rather than as a genuine aspect of an autopoietic system (*ibid*: 85–6).

2.5 Inputs, Outputs, and Perturbations

Both computing mechanisms and autopoietic systems turn out to be non-representational for much the same reason. This is essentially because of the way that they interact with the external world. Each kind of system is fully specified without any mention of its environment, in the sense that we can give a complete definition of a computing mechanism or autopoietic system that does not mention anything external to the system. This pre-empts the need for representational states, which typically represent something external to the system. Real-world systems do exist in an environment though, and so we must consider how they interact with that environment, and what makes this interaction non-representational.

As mentioned above, most actual computing mechanisms will include input and output components, without which they would be unable to interact with the external world. These components typically act as transducers, converting external stimuli into a format that is compatible with the mechanism's processors (i.e. strings of digits), and vice versa. To reiterate, though, the mechanism would still be computing in the absence of inputs or outputs, it would just not be of much use or interest to us, its users. Even if we interpret these inputs and outputs as representational, this need not carry over into the mechanism itself, which will continue to function regardless of our interpretation (see Dewhurst 2014: sec. 3).

Autopoietic systems are comparable in the sense that whilst they might not require anything external, they do nonetheless exist in an environment and will be influenced by that environment. Maturana & Varela call these environmental influences "deformations" and "perturbations", and note that they are indistinguishable from internal influences, at least insofar as the dynamics of the system are concerned (1980: 98). What happens is that the system's homeostasis is interrupted by an event (either external or internal), and the system then responds to the event, either by compensating in some way that returns it to homeostasis, or undergoing more radical change that constitutes the creation of a new autopoietic unity (*ibid*: 99). At no point does the system treat an influence as being external rather than internal, thus preserving the status of the system as non-representational. Whilst for practical purposes the system might sometimes require a way of distinguishing internal from external stimuli, this could take the form of a purely functional marker, and would not constitute the introduction of representation into the system.[4]

[4]I thank Paul Bello for bringing this last point to my attention.

Maturana & Varela also specify, "in terms of their functional organisation living organisms do not have inputs and outputs" (1980: 51). Here I take them to be referring to inputs and outputs in a classical sense, i.e. as processes that transfer representational content (see Piccinini 2012: sec. 2.3). The input and output components of a computing mechanism must differ from this classical sense if they are going to remain non-representational. Maturana acknowledges that an autopoietic system has "sensory and effector surfaces", which we could think of as comparable with inputs and outputs, but these do not preclude functional closure, because "the environment [...] acts only as an intervening element through which the effector and sensory [surfaces] interact completing the closure of the system" (Maturana 1975: 318). An autopoietic system is functionally closed because whilst it might respond to an external stimulus (i.e. a perturbation or deformation), the fact that the stimulus is external does not differentiate it in any way from an internal stimulus located at the sensory surface. It is, in effect, treated as a spontaneous internal event.

We should think of the input and output components of a computing mechanism in much the same way. From our perspective, looking in to the mechanism from the outside, they appear to operate in the classical sense that Maturana & Varela rule out. However, from an imagined internal perspective things look quite different. An input component simply produces strings of digits, whilst an output component consumes them. Taken together as a pair, they bear a strong resemblance to any other single processor, which consumes and produces strings of digits in much the same way. The only difference is that the processor is in this case constituted by the external world, which mediates the transformation from output string to input string. Functional closure is preserved, as the computing mechanism, like the autopoietic system, is functionally isolated from the external world. Thus, just the same as in an autopoietic or enactive system, input and output are "co-dependent aspects of a single circular process" (Villalobos and Ward 2014: 5), and there is no point at which representation can enter the picture.

2.6 Thinking Outside the Box

It is perhaps an unfortunate historical accident that computational and autopoietic/enactive approaches to the study of cognition came to be so diametrically opposed to one another. Both had foundations in early cybernetics, and only really began to drift apart after WWII. By the time Maturana & Varela published *Autopoiesis and Cognition* in 1980[5] the two traditions had been separated for several decades, and the computational approach to cognitive science had become dominant. It is in this context that autopoietic theory, and the enactivist tradition

[5]Maturana had published the first half of the book separately in 1972, under the title *Autopoiesis: The Organization of the Living*, and the second half was published a year later.

that it contributed to, is seen as a radical alternative to mainstream representational theories of mind. However, if the mechanistic account is correct, then there is nothing essentially representational about computation, and this divide between the traditions becomes somewhat weaker, perhaps even allowing for complete reconciliation.

This reconciliation would force the computational approach to reconsider the relationship between the brain and its environment. Whilst I have focused on the unity and closure of autopoietic systems, they are also fundamentally involved in their environment, as the later development of enactivism makes clear. Maturana describes this as a consequence of an autopoietic system being unable to distinguish between internal and external perturbations. For the system "there is no inside or outside", meaning that autopoiesis becomes inherently world involving (Maturana 2003: 99).

Here there is an important lesson for the computational theorist – computing mechanisms cannot be all that there is to cognitive science. The nervous system might well be computational, but it is also part of a situated organism, and a theory of cognition that ignores this will fail to capture the full complexity of its target domain. There is something missing from the picture – the world itself – that is restored when we take into account the role of worldly interaction in cognitive activity, understood as a mediating factor between output and input components.

On the other hand, enactivism has sometimes been criticised for failing to acknowledge the important role that the brain plays in cognition, or else failing to give a full account of what it is that the brain contributes to cognitive activity. For better or for worse, contemporary neuroscience is primarily computational, and accepting a non-representational account of computation could allow enactivism to become better integrated with mainstream empirical research. There seems to be no fundamental reason why a synthesis of the mechanistic account with autopoietic/enactive theory could not contribute to a fuller explanation of cognitive phenomena.

2.7 Conclusion

I have shown how an analogy can be drawn between computing mechanisms and autopoietic systems, focusing on the status of representations in both kinds of system. This analogy helps clarify the mechanistic account of computation by demonstrating that a computing mechanism treats paired input/output components as equivalent to processing components, thus preserving functional closure and pre-empting the need for representation. I have also suggested that the analogy might facilitate reconciliation between computationalism and enactivism, and that this would be beneficial to the study of cognition.

What I have not done is make any serious attempt to motivate why I think such reconciliation would be beneficial to both parties. My full thoughts on this must be saved for another occasion, but in brief I think that computation can offer enactivism

and autopoietic theory a convincing mechanistic base, and that in return enactivism and autopoietic theory can help computation escape from the metaphysical baggage that representationalism has burdened the received view with. In addition, paying attention to enactivism and autopoietic theory might help the development of non-traditional models of "sui generis neural computation", as hinted at by Piccinini and Bahar (2013), and considered more explicitly by Friston (2013).

I have also not considered the many dis-analogies between computing mechanisms and autopoietic systems, partly in the interest of space but mostly because they are not relevant to the claims that I am making. Computing mechanisms may not be identical with autopoietic systems, but proving that was never my intention. Rather I think it is unlikely that either approach will by itself fully explain cognition, and by comparing the two I hope to have gestured toward a synthesis that might further our understanding.

Acknowledgements I would like to thank Dave Ward, Paul Bello, Stefano Franchi, Mario Villalobos, four anonymous reviewers, and several members of the audience at IACAP 2014 for their helpful comments and suggestions.

References

Bechtel, W. (2005). Mental mechanisms: What are the operations? In *Proceedings of the 27th annual meeting of the Cognitive Science Society* (pp. 208–213). New Jersey: Lawrence Erlbaum Associates.

Craver, C., & Bechtel, W. (2006). Mechanism. In N. Sarkar & N. Pfeifer (Eds.), *Philosophy of science: An enyclopedia* (pp. 469–478). New York: Routledge.

Craver, C., & Darden, L. (2005). Introduction. *Studies in History and Philosophy of Biological and Biomedical Science, 36*, 233–244.

Dewhurst, J. (2014). Rejecting the received view: Representation, computation, and observer relativity. In *Proceedings of AISB 50*. http://www.doc.gold.ac.uk/aisb50/AISB50-S03/AISB50-S3-Dewhurst-paper.pdf

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

Egan, F. (1995). Computation and content. *Philosophical Review, 104*, 181–204.

Fodor, J. (1981). *Representations*. Cambridge: MIT Press.

Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface, 10*, 20130475.

Hurley, S. (1998). *Consciousness in action*. Cambridge: Harvard University Press.

Hutto, D., & Myin, E. (2013). *Radicalizing enactivism*. Cambridge: MIT Press.

Maturana, H. (1975). The organization of the living: A theory of the living organization. *International Journal of Man-Machine Studies, 7*(3), 313–332.

Maturana, H. (2003). The biological foundations of self-consciousness and the physical domain of existence. In N. Luhmann, H. Maturana, M. Namiki, V. Redder, & F. Varela (Eds.), *Beobachter: Convergenz der Erkenntnistheorien?* (pp. 47–117). München: Wilhelm Fink Verlag.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition*. London: Reidel.

Millikan, R. (1995). Pushmi-pullyu representations. In J. Tomberlin (Ed.), *Philosophical perspectives 9: AI, connectionism, and philosophical psychology* (pp. 185–200). Atascadero: Ridgeview Publishig Company.

Mingers, J. (1989). An introduction to autopoiesis. *Systems Practice, 2*(2), 159–180.

Noë, A. (2004). *Action in perception*. Cambridge: MIT Press.

Piccinini, G. (2004). Functionalism, computationalism, and mental contents. *Canadian Journal of Philosophy, 34*(3), 375–410.
Piccinini, G. (2007). Computing mechanisms. *Philosophy of Science, 74*, 501–526.
Piccinini, G. (2008). Computation without representation. *Philosophical Studies, 137*, 205–241.
Piccinini, G. (2012). Computation in physical systems. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Fall 2012 Edition).* http://plato.stanford.edu/archives/fall2012/entries/computation-physicalsystems/
Piccinini, G., & Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive Science, 34*, 453–488.
Ramsey, W. (2007). *Representation reconsidered*. Cambridge: Cambridge Universtiy Press.
Sprevak, M. (2010). Computation, individuation, and the received view on representation. *Studies in History and Philosophy of Science, 41*, 260–270.
Thompson, E. (2007). *Mind in life*. Cambridge: MIT Press.
Varela, F. (1991). Organism: A meshwork of selfless selves. In A. I. Tauber (Ed.), *Organism and the origin of self* (pp. 79–107). Dordrecht: Kluwer Academic Publishers.
Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge: MIT Press.
Villalobos, M., & Ward, D. (2014). Living systems: Autonomy, autopoiesis and enaction. *Philosophy of Technology*, online first. doi:10.1007/s13347-014-0154-y.
Weber, A., & Varela, F. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences, 1*, 97–125.