1. Introduction

As deterministic and discretely operating machines, modern-day digital computers seem to contradict the foundational assumption of constructivism. An epistemology that stresses the skeptical argument that an "unconditioned first," such as an unmoved mover or an unobserved reality, is no viable base for analytical science (Foerster 1976; Glasersfeld 1995; Luhmann 1995) does not appear to conform to a technology that inevitably is subjected to the grounding problem (Harnad 1990). This problem arises when whatever is processed with the help of that technology needs to resort to some sort of externally given "first," to a starting value or an initial condition, that is usually defined by a programmer who herself is not part of the computation.

Indeed, so I will argue, computers in their current form do not meet the requirements of modeling truly self-referential end-directed systems. After all, they are just analytical tools apt to modeling abstracted cutouts of what we perceive as world. However, they enable a comprehension of the interaction of non-linear dynamics to an extent that sheds light on possibilities of how to explain scientifically several so-far-unanswered questions concerning the constituencies of life, the functioning of cognition and the emergence of mentality.

The paper is organized as follows: Section 2 will briefly lay out the basic assumptions and implications of constructivism. Furthermore, it will discuss some constraints of digital computation, as well as outlining the general concept of computation. This concept is not restricted to digital data processing on machines but comprises the possibility of what has been envisioned as "pancomputationalism" by Konrad Zuse or Edward Fredkin. Section 3 will elaborate on the observable as a central constructivist conception. It will review insights from second-order cybernetics, the algebra of forms and the mathematical conception of attractors. Subsequently, Section 4 will introduce the so called Bayesian brain approach of contemporary robotics as it builds on assumptions similar to those of second-order cybernetics. In order to clarify the constructivist implications of this approach, Section 5 will first review aspects of the debates about the term emergence and will outline Mark Bedau's suggestion for a definition of this term in terms of computer-based simulation. In regard to objections to the concept of emergence that concern basic preconditions of digital computation, Sections 5.2 and 5.3 will then address the problem of downward causation and demonstrate its simulatability with a version of Joshua Epstein's demographic prisoner's dilemma. Finally, we will briefly summarize the arguments that seem to support the case of constructivism when viewed through the lens of computation.

2. Constructivism and computation

Drawing on suggestions by Heinz von Foerster, Ernst von Glasersfeld, Niklas Luhmann and others, one might summarize the central constructivist thesis with a statement of the kind: "what is perceived as reality is the construction of an observer." If one subsequently agrees that this observer itself should be regarded as part of reality, one is forced to conclude that "the observer is a construction of an observer." From this, ensues a circularity that irritates many natural scientists. Up to now, many of them have seemed to find it less unsettling to believe in a reality that exists independently of whether it is observed or not.

Indeed, on first sight, constructivism seems irritating and contradictory to intuition. Each formulation of its tenet seems to necessitate specifications that lead straightaway into what the classics have
called – and repudiated as – a vicious circle. Objections are manifold and have been put forth in respect of nearly all its historical variants. Objections range from that of classical solipsism as well as Descartes’ dream argument, which began continental subject philosophy, to the debates on constructivism’s self-implying relativity (Gillett 1998). They also include the infamous “brain in a vat” hypothesis (Dancy 1985: 10) and related speculations such as that of Frank Tipler (1995) about us and our universe being embedded in a computer simulation that is running on a computer in another universe. The conclusion that there is no proof that this simulation is again not just a part of yet another more universal simulation is either self-referential and therefore recursively circular – “it is turtles all the way down” – or it entails the old philosophical quest for a “first,” in this case for an “unsimulated simulator.”

Since digital computers operate deterministically and discretely and are therefore often seen as the epitome of classical logic, they seem to raise this issue in their own right. Their calculations have to “hit bottom” somewhere, otherwise they would not work in the way we expect them to. On the other hand, however, digital computers – as we shall see – enable a clear and comprehensible conception of recursions and the possibility of detailed investigations into their dynamics and consequences. Only with their availability could it become clear that non-linear dynamical interactions that in a certain way seem “bottomless” can run up to sufficiently stable orders that provide a footing for next-order interactions, which in their turn again might seem “bottomless” but are able to create realities sui generis in their own right, and so on. In short, I will argue that the digital computer opened up analytical access to the behavior of complex non-linear dynamics that hitherto simply escaped observation. Only with the computer did this behavior become accessible to the extent to which it is now starting to change analytical thinking.

3. The observer

In this respect, the above mentioned speculation of Tipler’s simulated universe meets with another seemingly irritating assumption that was first expressed by computer pioneer Zuse (1967) in his paper on “calculating space” (in Zenil 2012) and was recently picked up by researchers such as Jürgen Schmidhuber (1997), Stephen Wolfram (2002), Edward Fredkin (2003) and others. It consists of the assumption that natural processes are forms of computation themselves, and that our universe can be regarded as being one huge Cellular Automaton running since the “beginning of time” and constantly constructing new realms of order that, in their turn again, interact in complex and momentous ways.

Without doubt, what is meanwhile discussed under terms such as “pancomputationalism” or the “Zuse-Fredkin-Thesis” implies reductionism. This is the assumption that what we observe as realities sui generis (the best examples may be our consciousness, our self, or living systems, etc.) can be reduced to biology, which can be reduced to chemistry, which can be reduced to physics, which can eventually be reduced to computation of information at a most fundamental level. However, exactly at this fundamental (but not substantial in the sense of ontology) level, I will argue, the close connection between constructivism and computer-based modeling can become clear. This is because simulation unmistakably (since experimentally repeatable, as we shall see) illustrates how realities sui generis (a.k.a. systems) can emerge from lower-level interactions that do not necessarily carry their properties (cf. Section 5.1 below).

In this respect, the Zuse-Fredkin-Thesis has the interesting implication that what is now being done in terms of computer-based modeling and simulation could be seen as a sort of “algorithmic compression” of the universe copying what could be called the strategy of nature to enclose and separate (differentiate) temporary “world representations,” i.e., systems, which in their own turn are sufficiently complex to generate next-order worlds again, and so on. The thereby implied duality of order generation and representation seems to be graspable in terms of the relation of an observed to its observer. This relation is at its core, as has been suggested, a recursive circular relation.

Since according to constructivism, whatever is perceived as reality is not conceivable without being observed as such, a central conception of constructivism is the observer (Foerster 1976). To a constructivist, an independent, i.e., unobserved, reality simply does not exist. Reality is always to an observer. And as we have said, this observer has to be conceived as being observed itself (see also Füllsack 2012b).

The maybe most fundamental conception for bootstrapping such an observer seems to be the distinction/indication duality of George Spencer-Brown (1969), in which observation is seen as a formal duality of distinguishing something and indicating one of the two distinguished sides. The act of indication thereby refers to a basic binary choice for one of the two sides. This definition is formal in the sense that it comprises an air-conditioning system that observes a rising room temperature by distinguishing “too hot” from “normal” temperature and indicating “too hot” with an on-signal to the cooler, in the same way as a scientist observes the behavior of an object by delimiting it into the framework of a controlled laboratory situation. The circularity of this conception arises from the fact that observations always build on preceding observations, which have to be presupposed but cannot be observed in the current act of observation, thereby generating an “unmarked space” in each observation. Just as the air-con presupposes temperature (as distinguished from any other quality a room might possess), the scientist builds her activities on the need for tested scientific insights (as opposed to beliefs, for instance). Both these presupposed and thus unmarked spaces can only be dissolved in further observation, that is, in observation of observation or second-order observation, which faces the same consequence in its own turn. Since this is true for any observation, the distinction/indication duality implies a potentially infinite recurrence of observations that have no beginning and – if conditions are viable – no end. Each observation is subjected to observation itself, implying an ongoing interchange and interaction of observed and observer.

1 | This refers to the by now well-known joke recounted by Stephen Hawking in A Brief History of Time (Hawking 1988: 1).
Drawing on Jean Piaget's conception of cognitive development via ongoing subject-world interactions (to which I come back below), von Foerster (1976) pointed to the mathematical aspects of respective recursions in which each output is used as its next-order input, thereby implying a unity of operator and operand. Some of these recursions tend to run up to attractors, or, as von Foerster called them, eigenvalues or eigenforms, that is, to invariant subsets of the phase space of interaction. For instance, the eigenvalue (in this case the fixed-point attractor) of the operation \( \text{op}(x) = \text{divide } x \text{ by } 2 \) and add 1" is an \( x \) such that \( x = x/2 + 1 \), which is 2. In iteration, the solution for this recursion is found by choosing any number for \( x \) and subsequently applying \( \text{op} \) repeatedly to its own results, just as an assumed observer would use its observation in time \( t \) as a precondition for its observation in time \( t+1 \). The essential aspect thereby is the fact that this iteration runs up to the value of 2 regardless of what initial value is chosen. So 2 is the eigenvalue of this recursion irrespective of its first value. In other words, the initial value is irrelevant. The operation heads for 2 on its own. One might interpret this circumstance (which, by the way, is a reason to conduct systems sciences as a distinct discipline that investigates the specific logic of independent spheres of eigenbehavior) as an operation with no need for any "ontological base" or any "external reality" in order to generate spatially and temporarily stable conditions. The order of this operation emerges from noise.\(^2\)

Such orders are not restricted to simple fixed point attractors. Referring to the considerations of von Foerster and Spencer-Brown, Louis Kauffman (2005, 2009) introduced a rich collection of recursions leading to interesting and surprising eigenvalues. One impressive example, for instance, is John Conway’s "audio-active sequence" resulting from an iterated oscillation between numerical and verbal descriptions of a number that after a couple of recursions runs up to the eigenvalue of a reappearing numerical triplet. Another nice example is the form of the division of the golden rectangle, as discussed in Kauffman & Varela (1980). With regard to the equation \( T(x) = -x^4 \) (respectively \( T(x^2) = -1 \), having a complex solution), Kauffman illustrates how the eigenbehavior of its recursion

\[
T(T(T(\ldots))) = \cdots(-(-\cdots-1)-1)-1
\]

generates a wave train of the form \(-1, +1, -1, +1, \ldots\) which has points in common with Spencer-Brown’s (1969) assumption about the generation of "mathematical time."

Still more complex examples of eigenforms can be seen in what, in the mathematics of non-linear dynamics (Strogatz 1994), have been called "strange attractors," with the Lorenz butterfly possibly providing the most famous example. Being generated by the recursive iteration of several interacting dynamics and thereby mathematically depending on the solution of coupled differential equations, an analytical investigation into these phenomena remained severely restricted, if not impossible, in pre-computer times. Although some far-sighted scientists such as Henry Poincare anticipated some of their specifics, the detailed constitution and the terms of emergence of these systems became assessable only with the advent of the digital computer. Some rather simple dynamical interactions such as the Lorenz system were investigated relatively early in the digital age. However, the full significance of higher-order phenomena emerging from lower-level non-linear interactions only seemed to become graspable recently, with the rapid growth of CPU power and memory in digital machines allowing for simulations with interacting dynamics in the scale of several thousands. I will come back to the issue of emergence later.

As mentioned before, von Foerster resorted to the mathematical concept of attractors with respect to suggestions by Jean Piaget about object invariance emerging in the course of an iterated interplay of observation and movement of a person trying to come to terms with its environment. Piaget considered this sensorimotor interaction as being circular and recursive in the way that each observation (\( \text{obs} \)) induces a subsequent coordinative movement (\( \text{coord} \)), which in its turn changes the observational viewpoint, albeit slightly, and thereby again necessitates an observation that might induce movement and so on. When formalized, this recursive interaction can be denoted as an \( n \)-length sequence of \( \text{obs} \)- and \( \text{coord} \)-functions working on each other and thereby generating stable eigenforms in the above sense that might eventually be perceived "internally" as "objects."

Starting with an assumed \( \text{obs}_0 \), this recursion might be depicted as follows:

\[
\begin{align*}
\text{obs}_1 &= \text{coord}(\text{obs}_0) \\
\text{obs}_2 &= \text{coord}(\text{obs}_1) = \text{coord}(\text{coord}(\text{obs}_0)) \\
\text{obs}_3 &= \text{coord}(\text{obs}_2) = \text{coord}(\text{coord}(\text{coord}(\text{obs}_0))) \\
& \quad \quad \vdots \\
\text{obs}_n &= \text{coord}(\text{coord}(\text{coord}(\text{coord}(\text{coord}(\ldots(\text{obs}_0)))))) \\
\end{align*}
\]

However, as in von Foerster’s and Kauffman’s examples, this interaction renders its initial value irrelevant when \( n \) approaches infinity. It is not important where it “hits bottom,” it generates invariance by itself.

http://www.univie.ac.at/constructivism/journal/9/1/007fuellsack
Currently, these considerations of Piaget seem to find confirmation in the way cognitive activities are simulated and implemented in AI and robotics (Der & Martius 2012). The so-called Bayesian brain approach (Knill & Pouget 2004; Doya et al. 2007) combines several techniques that build on a comparable sensorimotor interplay. Central to it are computational techniques known as recursive Bayesian estimation and predictive coding, as used in signal processing.4

Recursive Bayesian estimation (Robbins 1956) starts out with a system having a “belief” in the form of a prior probability distribution about the states of its world. This “belief” in the form of an internally maintained representation or a “world model” might be pre-defined (by an engineer) or just randomly generated at startup. It might hence be far from what in a conventional sense could be called a “veridical” representation or a consistent model. When initiated, the system enters into a recursive interplay of observation and movement, as mentioned above. In the course of this interplay, this internal representation (the prior) is iteratively updated with new sensory inputs, thereby at each step computing a posterior probability distribution for the states of the world. In the next iteration, this probability distribution is treated as a new prior, i.e., a new representation again, which in its turn then serves as a prior for a new posterior, and so on.

The terms updating and representation, however, have to be understood in a particular way in this case. Although widely used in machine learning and robotics, it is important to note that to the system, this iterated updating is not an aim in the sense that the term learning might imply. The system is not actively refining its world model by reacting to new sensory inputs. On the contrary, to the system, each iterated representation reflects what the machine “knows” about its world at the current moment of time. This knowledge governs the system’s observational and motoric activities and allows for certain predictions of what the world is like. The system tries to fulfill these predictions, and therefore uses its motoric possibilities. One might say that it tries to “reconfirm” its world knowledge by using motor actions, rather than trying to refine or enhance its knowledge. In moving, it tries to match sensory inputs with what it can currently predict about its world.

In order to detail this aspect of reconfirmation, it is necessary to bring to mind that in this process the world itself is off-limits to the system. All it knows are its own states, which might be constantly flowing and changing as the system iterates between observation and movement (Rijke 1999; Der 2008). Still, these states represent the system’s world, or more to the point, its world construction. Thereby, the system has no notion of a deficient representation that has yet to be updated and refined. All there is is its current representation in each moment of time. If there should be a mismatch of a sensory input and the representation, this is not a valuable new insight or new evidence about the world to the system. It is just an “irritation” (Luhmann 1995) that has to be integrated into the system’s knowledge in order to reconfirm the representation it provides. To move hence – that is, to perform a motor activity – is one of the system’s ways of attempting to match sensory inputs with current knowledge (and there might be several other ways on higher-order levels). In regard to this motoric, one might say that the system actively tries to fulfill its predictions. This notion of prediction fulfillment can be further refined in terms of an open loop technique that in signal processing is called predictive coding (Shi & Sun 1999). It consists of compressing data by encoding only the “unexpected” variations in an image, that is, those cases where, for instance, the actual value of a pixel departs from what can be predicted from neighboring pixels. What is encoded in this technique is not the complete actual information of an image, but just the difference between a current signal and the one that can be predicted from surrounding signals. This difference is called the “prediction error” and is used in various forms of lossless audio or video recording to afford major savings in bandwidth. In machine learning and robotics, it is successfully deployed to account for how signals are processed in the visual cortex (Rao & Ballard 1999) or how the retina economizes on bandwidth and singles out what is most newsworthy in an incoming signal (Hosoya, Baccus & Meister 2005). Currently, it is thought to provide a general model for the way the human brain comes to terms with a continually changing world (Friston 2009; Clark 2012).

The essential point in this prediction reconfirmation is hence that the system can be seen as “actively trying to explain away” the prediction error (Clark 2012). It behaves “conservatively” in the sense that it does not look for new data to update its world representation. It rather tries to fulfill its expectations – and therefore deploys its motoric, which closes the loop. A mismatch of its world representation and a sensory input makes it move in order to clear the error away, as could be shown in experiments with machines actively moving their sensors in order to generate the sensory consequences that they expected from their world models (Friston, Daunizeau & Kiebel 2009).

Only if moving does not succeed in matching signal and representation does the Bayesian update take effect and a new posterior is computed. However, this update follows Bayes’ rule and therefore depends on the prior, that is, on the preceding state of the system, i.e., on its “knowledge.” The system integrates an irritation from a mismatch of representation and sensory input on the basis of its own knowledge. This means it integrates it in its own particular

4 In terms of control theory’s differentiation between open and closed loop learning techniques, recursive Bayesian estimation could be regarded as representing a closed loop process, implying that each operation’s output is fed back into the system’s next iteration. Predictive coding, on the other hand, is not based on feedback by itself and thus could be seen as an open loop technique. The Bayesian brain approach integrates these two learning techniques, making the whole process feedback-driven and thus a closed loop. On the background of constructivist epistemology, however, this differentiation (and speaking about open loops in particular) might be misleading since it refers to input/output processes that imply something like an “inside” as distinguished from a (possibly independent) “outside.” The Bayesian brain approach seems constructivistically interesting since it revokes the inside/outside differentiation, thereby implying a sort of concurrency – a “synchronous asynchrony” (Füllsack 2011b) – of causes and effects, that is, of environment and system or of observed and observer, as I will discuss in more detail in the next section.
way. In this sense, it does not react to any “independently existing reality” or to an informational map of this. It reacts to a “prediction error” that is basically an “interpretation” of an irritation generated on the basis of the system’s current possibilities. In this sense, the system sees what it can see, and it does not see what it cannot see. The system constructs its world on the basis of its own unique possibilities.

5. The causation issue

However, as this “action-oriented predictive processing,” as it is called by Clark (2012), is currently primarily considered in regard to endeavors to ground a unified science of mind, brain and action by way of implementing it in individual robots5, it might still at times inspire associations with the realist’s picture of an individual brain – or an individual system – being exposed to an “external” environment as its distinct source of incoming data. Even though not directly conveyed but processed through internal states, this data then could be seen as existing independently of an observer. After all, there is something by which the robot (the system) is irritated, a

see talking Heads experiment by Luc Steels (1999). See also Füllsack (2012a).

5.1 Emergence

In logic, causation is conceived as unidirectional, that is, as working from cause to effect and not the other way around. Computers are deterministic and discretely operating machines that comply with logic, as do the mathematical recursions we considered in Section 2. Each of these recursions builds on the results of its precursor in the preceding iteration and hence follows a strict chronological order. Algorithms can perform loops, of course, and at times quite complex ones, and they can be executed in parallel on several CPUs. However, even in these cases the basic order of calculation remains strictly sequential. Nevertheless, recursive calculations can run up to quite complex constructs or forms at times, as discussed in Section 2 on the example of attractors and eigenforms.

However, these eigenforms, due to their relative stability, are prone to be perceived in their being (i.e., ontologically) rather than in their becoming (i.e., ontogenetically). Intuition – or as we shall shortly insinuate below, the limited computational power of the observer – therefore caused and still causes European epistemology to focus on substances and objects instead of the processes that bring them about. In particular, in pre-computer times the focus was bound to remain on the results or consequences of what now can be pictured as interacting recursive operations, since these were simply not easily disentangled. It is an arduous task indeed to calculate by hand even relatively simple coupled differential equations, such as those responsible for the Lorenz attractor.

Scientific analysis therefore remained restricted. In the 19th century, science suspected several phenomena to be caused by an interaction of underlying components, but the details of how the properties of these phenomena were brought about remained unclear. The fluidity or transparency of water, for instance, remained enigmatic since it was known that water consists of hydrogen and oxygen, but also that both of these substances show none of these qualities on their own. In order to refer to the fact that higher-order properties might arise from the interaction of lower-level components lacking these properties, such phenomena were called emergent (cf. Kim 1993; Holland 1998; Clayton & Davies 2006). In being associated with the Aristotelian notion of a whole being more than the sum of its parts, they entailed somehow vague and at times even unscientific connotations. It remained unclear whether something – and if so, what – was missing in the explanations this term provided. When 20th-century science managed to explain some of the properties considered as emergent – for instance, with the help of quantum-mechanics or biochemistry – the term was discarded from natural sciences. In some other disciplines it remained in use, but was confined to denote such hard-to-grasp phenomena as consciousness, life or the disputed “qualia” of phenomenological philosophy.

In the 1960s however, with the rising interest in self-organizing processes and in particular with the advent of computer science, the discussions about emergence resurfaced. First in the context of research on cellular automata (CA), and later in what was to become agent-based modeling, the term proved to be useful for denoting the phenomenon of pattern formation. This phenomenon was observed, for instance, in the case of a deterministic rule set defining the interrelation of just three neighboring cells but generating reappearing patterns spanning over 20 neighboring cells, as is the case in the rule-22 automaton of the basic one-dimensional CAs, as investigated by Stephen Wolfram (2002, cf. Figure 1, left and middle image). Similar global (or macro-) effects of just local (or micro-) interactions could be shown with regard to segregation, traffic jams, patterns on animal skins and furs, the synchronization of mov-
CoMPutAtIonAl ConCePts In RAdICAl ConstRuCtIvIsM

emergent properties can be considered another occasion. To current space and topic restrictions I save for another occasion. The disputes about emergence hence seem to have gained from the possibilities of digital computation. But what does this mean for the above-mentioned environmental issue and the question of whether digital computers allow a comprehension of co-evolution and what has been called downward causation?

5.2 Downward causation

Currently, respective insights seem to provide fertile ground for endeavors such as that of Terry Deacon (2011) to unify explanations for what are conventionally distinguished as the topics of natural sciences and of humanities, that is, physical processes that do not seem to have an aim or "ententional," as Deacon calls them) teleodynamic processes in terms of an emergence from the interaction of dynamics on the preceding level of order. This attempt clearly builds on the possibility to conceive emergence as a thoroughly grounded scientific conception that does not resort to any hidden or obscure causes.6

Deacon emphasizes the role of computer technology in this development, too. Interestingly, however, he repeatedly objects to the possibility that the observer might play a role in what appears as emergent (cf. Deacon 2012: 6, 74). This point deserves a thorough discussion, which due to current space and topic restrictions I save for another occasion.

6 | Deacon emphasizes the role of computer technology in this development, too. Interestingly, however, he repeatedly objects to the possibility that the observer might play a role in what appears as emergent (cf. Deacon 2012: 6, 74). This point deserves a thorough discussion, which due to current space and topic restrictions I save for another occasion.

7 | Campbell seems to have coined the phrase downward causation. His epistemology, however, is as is that of his ally Karl Popper – explicitly based on realism and thus opposed to constructivism.

<table>
<thead>
<tr>
<th>left neighbor</th>
<th>right neighbor</th>
<th>( x_{t} )</th>
<th>( x_{t+1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 1 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 1 0 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 1 1 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0 0 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0 1 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1 1 0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Left: Look-up table for Wolfram’s "rule 22," comprising a "local" interaction of just three neighboring cells. Center: Pattern generated by this rule. Right: Cone snail.
seen as a collective product of interacting individuals, which, as it evolves, definitely feeds back onto the cognitive development of these individuals. The question that bothered philosophers of science, however, was whether this kind of mutual causation fits to the concept of emergence. In a very condensed form, the objection of people such as Jaegwon Kim (1993) focused on the question of how the influence of downward causation can take effect if, according to reductionism, realism and European substance philosophy, the entities causing the emergent phenomenon are elements that, at least at the very end of a possibly long causal chain, have to be considered as substances of last resort, as atomistic particles that are no longer divisible. Translated to the realm of digital computation the problem consists in the way the emergent phenomenon, the macro-level of order, can alter the elements causing it if these elements can only be the discrete initial values that computer technology needs for a computation to be started.

Concerning the philosophical problem, the argument was rejected in respect to insights from quantum theory that gave reason to assume that an ultimate atomic particle or anything that can be regarded as substance in the sense of classical European philosophy simply does not exist. Quantum theory dissolved this base at the bottom. At the lowest level of scale, there are only quantum fields that are not graspable other than statistically and dynamically, usually in the form of a wave function. Therefore, what traditional philosophy calls substance or particle has to be seen as itself made up. The lower-level elements of an emergent phenomenon are emergent themselves, with their elements again being considered emergent and so on. It is turtles all the way down.

Hence, if the conception of quantum fields and their dynamical constituency holds water, one would have to dismiss the classical whole/part dichotomy and confine oneself to a processual world. The question in this context, however, is of how computer technology, and in particular simulation, relates to this processuality. Can phenomena such as downward causation, or in respect of the Bayesian brain robot of Section 4, the coevolution of system and environment, be conceived in simulation? And if so, would this allow generalizations of concepts such as environment or embodiment that are considered indispensable in contemporary robotics? The following section will introduce a model meant to demonstrate aspects of this possibility.

### 5.3 Cooperation

One famous example of a case where computer-based simulation could increase understanding significantly is the emergence of (game-theoretic) cooperation. Not least with the help of simulation, it could be shown that cooperation might emerge not in spite of, but on the contrary, as a consequence of selfish agents acting to optimize their individual payoffs (cf. Axelrod 1984, 1997; Nowak & Sigmund 1998). In particular, a proposal by Epstein (1998, 2006) on how to transpose the famous prisoner’s dilemma tournament of Bob Axelrod (1984, 1997) from a temporal to a spatial dimension seems to illustrate the issue in question.

In this so-called demographic prisoner’s dilemma (DPD), Epstein, like Axelrod before him, suggested confronting autonomously acting computer-generated agents repeatedly with each other in prisoner’s dilemma-interactions. A prisoner’s dilemma (PD) (cf. Rapoport & Chammah 1965) is defined as a symmetric two-player normal form game in which the payoff of the action of one player depends on the choice of action of the other player. The particular PD rules define the payoff for one-sided defection (“temptation,” T) to be higher than the “reward” (R) for mutual cooperation, which in its turn has to be higher than the “punishment” (P) for mutual defection, which again has to be higher than one-sided cooperation (“sucker’s payoff,” S). In short: T>R>P>S. In PDs, therefore, defection (i.e., non-cooperation) is the dominant strategy. It is definitely a player’s safe option if the opponent should defect as well, and it still yields a higher payoff than anything else if the opponent should cooperate. Since this is true for both players alike, the question arises of how, under these conditions of double contingency (Parsons & Shil 1951; Luhmann 1995: 105), cooperation can emerge as a stably established (which means not just a coincidental) behavior. Bob Axelrod answered this question in regard to the possibility of repeating PD confrontations (IPD for “iterated PD”) and therewith iteratively collecting information about the most likely behavior of one’s opponent. The expectation of cooperative behavior – the “shadow of the future” as Axelrod called it – might then render mutual cooperation and its payoff R more attractive than defecting (and its double-sided payoff P).

Unlike Axelrod’s famous tournament, Epstein’s DPD allows for a stochastic confrontation of players. At startup, agents are randomly dispersed on a 30×30 torus grid. In each step of the game (in each iteration), they move to an empty grid patch within their vision and play a PD against all Von-Neumann neighbors they encounter, that is, against all neighbors on the patches immediately to the North, East, South and West of the agent’s own field. They thereby follow an innate strategy. Either they cooperate or they defect. Agents are hard-wired in this respect.

Again different from Axelrod’s tournament, Epstein’s DPD allows negative payoffs, as specified in Table 1.

<table>
<thead>
<tr>
<th>Agent A</th>
<th>cooperation</th>
<th>defection</th>
</tr>
</thead>
<tbody>
<tr>
<td>cooperation</td>
<td>R=5, S=5</td>
<td>S=–6, T=6</td>
</tr>
<tr>
<td>defection</td>
<td>T=6, S=–6</td>
<td>P=–5, P=–5</td>
</tr>
</tbody>
</table>

Table 1: Payoffs in the demographic prisoner’s dilemma according to Epstein (2006: 201).
hood, an offspring is generated who inherits 6 points of its parent’s wealth plus its strategy. Additionally, agents are endowed with a common maximum age and initially are “born” with individual random ages in order to prevent periodic waves of deaths from old age.

When running the simulation, cooperation reliably emerges and solidly establishes, even if the initial percentage of cooperators is very low. A typical run results in a population of cooperators (white dots in the images in Figure 2) that is punctuated by some islands of defection (red dots), with an approximate ratio of five cooperators to one defector.

As said before, Epstein’s DPD transposes Axelrod’s IPD from a temporal dimension to a spatial one. What in time is accomplished by the memory of the agents, that is, by their memorized experiences of the behavior of their opponents, is accomplished in space by the possibility of reproduction. Payoffs here are seen as a sort of “fitness” that enhances the reproductive success of an agent. The fact that agents inherit the strategy of their parents and therewith, if successful, increase their proportion in the population can be seen as an (statistically dispersed) equivalent to the memory of the agents in Axelrod’s model. In a certain sense, thus, evolution in the DPD appears to be caused linearly: initially (on level \( n \)) there are – albeit only a few – cooperators and then (on level \( n \!+\! 1 \)) there is – little wonder – cooperation. One might interpret this as a kind of presupposition. Agents are substances of last resort. They are particles that might organize, but are not conceived as organized on their own.

True, in a variation of its DPD, Epstein considers mutation. With a certain probability, new-born agents do not reliably inherit the strategy of their parent but mutate to the opposite behavior. In this case, with the above-tabled payoffs, cooperation remains stable, even at high mutation rates, but it does not affect the population in the same way as with no mutation at all.

Mutation, however, in Epstein’s model does not seem to be much more than some kind of perturbation of an otherwise “normal” inheritance of strategies. There is still at least some percentage of cooperating agents “at first,” from whose interactions cooperation “then” can emerge as the predominant strategy. In regard to behavioral evolution, this does not seem to be overly plausible. Undetermined at “birth,” agents should develop stable behavioral habits only in the course of their interactions, depending on their environment.

If new-born agents are considered coincidentally determined, however – that is, if they are born with a mutation-probability of 100% – cooperation becomes transient. It crops up in the population, but maintains an erratic character. If, additionally, agents then are exposed to a somewhat “rougher” environment, that is, to payoff values of, for instance, \( T = 9, R = 2, P = -5, S = -7 \), cooperation no longer has any chance. After a few iterations, cooperators die out and with them, a few time-steps later, defectors are extinct as well.

Therefore, in a variation of Epstein’s DPD (that is vaguely reminiscent of a “nurture over nature” principle), I allowed an initial mutation-probability of 100% to develop, depending on the coincidental confrontations of the agents with one or the other strategy (cf. also Füllsack 2011b). In this version, the probability of an agent to adopt a strategy grows with the frequency of being confronted with this strategy. We might say it grows according to an agent’s coincidental environment, which in itself becomes less coincidental as the agents adopt their specific strategy. For this, the 100%-mutation probability of newborn agents gradually decreases with each interaction until the agent is eventually firmly inclined either to cooperate or to defect. In this way, agents adopt their strategies evolutionarily, so to speak, over several generations when in some part of their world coincidentally temporarily stable clusters form, be they of cooperators or defectors. As a consequence, new-born agents will primarily adopt the strategy prevailing in their environment, thereby shaping this environment by their own development, while being shaped by it.

---

*Figure 2: Simulation of the demographic prisoner’s dilemma according to Epstein (2006), “cooperators” white, “defectors” red. Left: Initial random distribution of 100 agents. Center: After 15 iterations. Right: After 50 iterations.*
In this scenario, hence, the emergence of cooperation does not only depend on the micro-level behavior of the agents. The emergence itself feeds back on the agents’ constitution as well and changes their probability of how to behave. There is not only a causality from the n − 1-order level of interaction (from "micro-motives") to the n-order level of the emergence of cooperation (to "macro-behavior"), but also a downward-causation from the n-order level back to the n − 1-order level. Or, in other words again, there is no fixed environment, no "external world" on one side from which data comes in that shapes agents’ behavior on the other side, but a bidirectional causality that shapes the behavior of agents, which, at the same time in its turn, determines the environmental influence. One might say that not only the agents, but the environment, too, is simulated in its becoming in this case. What is more, in order to analogize this DPD example to sensorimotor interaction as discussed in paragraphs 15 and 16 above, one might even regard the agents as artificially "embodying" cooperation. One could then maybe compare this example to robot simulations such as those done by Ralf Der and his colleagues,8 where embodying activities of robots are simulated in silicio, that is, in artificial environments. The important difference to the DPD example, however, is that in these simulations no feedback of bodies (robots) to the environment is considered. The environment (as well as some preconditions of the body of the robots) is simulated in its being and not in its becoming, whereas here we try to simulate an integrated process of co-evolution that does not presuppose a steady "firstness" of a particular environment in regard to a "secondness" of the emergent property – cooperating agents in this case. Instead it proposes an intermingled causality of mutually influencing dynamics (Füllsack 2011a). This circular causality, which cannot be analytically dismantled without digital computation in a classical sense, gave reason to try to summarize respective phenomena with the term "synchronous asynchrony" (cf. Füllsack 2011b). Nigel Gilbert (1995) describes similar aspects as "second order emergence" and Cristiano Castelfranchi (1998) speaks of an "immergence" with which the effects of emergent properties affect the "cognitive" apparatus of the individual agents whose interactions generate them.

When running this variant of the DPD, the rudimentary possibility of evolving mutation rates suffices to raise the probability of cooperation over time high enough for the population to prosper. Whenever in one part of the grid – in a "niche," so to speak – a small island of cooperators coincidentally persists long enough to make their mutation probabilities sink to a level at which at least some of their offspring are reliably born as cooperators, cooperation can get a hold and develop. Although it never reaches the levels of the no-mutation variant – its fluctuations show typical predator-prey dynamics – it establishes a safe enough footing not to die out.

The principle is not new to scientific attention. It has been discussed under

---

8 | Cf. Der’s homepage at http://www.informatik.uni-leipzig.de/~der/

---

http://www.univie.ac.at/constructivism/journal/9/1/007fuellsack
terms such as “co-evolution of mutually conditioning causes” (Jantsch 1980: 207), and it has been mentioned in respect to hyper-cycles (Eigen & Schuster 1977), autocatalysis (Kauffman 2000; Deacon 2011) and autopoiesis (Maturana & Varela 1987), and is currently being discussed as the eco-evo-effect (Pennisi 2012). It oriented research agendas such as connectionism (Bechtel & Abrahamsen 2002), systems theory (Luhmann 1995) and network theory (Newman 2010). The principle is definitely not easy to disentangle, and to follow each recursive step in its micro-causal details would be arduous indeed. As with emergence before, however, it becomes tangible in silicio and thereby opens up to analytical investigation.

6. Conclusion

It could be objected that the above model does not demonstrate “bottomlessness” in the strict philosophical sense. Like any computer-based model, it resorts to an initial value from which recursions are started, to a “first,” hence, that represents some kind of external reality. At best, therefore, it can be said that it models a cut-out of an overall process of interacting recursions, and this cut-out only transfers the environment/system issue that it tried to assess to another level at which it might crop up in new form.9 The model, hence, neither implies, let alone proves, that a substance of last resort or an unobserved reality does not exist after all.

As this is certainly true, it has to be stressed that it was not the objective of these considerations to refute realism. To some extent, I think, it will remain a matter of taste whether one prefers to base scientific insights on a circular and at times counter-intuitive conception or on “knowledge” about a reality that can never be known.

What I tried to show in this paper is that constructivism and computation do not contradict but, on the contrary, are able to enrich each other productively. Computer-based models such as the one on downward causation cannot prove constructivism, but they can open up respective constructions to detailed analysis. They allow for a step-by-step investigation of interactions that remained widely impenetrable prior to the age of digital computers. They thereby provide chances to “crawl the micro-causal web” of multiple non-linear interactions, that is, of interactions that without these techniques are just too complex to be accessed in their details. With this, they open up possibilities for generalizations that eventually might exceed the limited domains that single computer models can currently cover. And through this, they might create even more reasons to consider constructivism a viable philosophy.

9 | Referring to the pancomputationalism that was mentioned in Section 2, one might, of course, speculate here about the possibility of computer-based models modeling other computer-based models that again model other models, and so on.

Received: 28 September 2012
Accepted: 19 September 2013

Manfred Füllsack
is Professor of Systems Sciences at the University of Graz. His research includes: systems, complexity, networks, games and computational theory, work – its history, its sociology, its economy, and its philosophy – and computer-based simulations.

OF RELATED INTEREST GLEICHZEITIGE UNGLEICHZEITIGKEITEN
In this book, Manfred Füllsack conveys insights from disciplines like cybernetics, systems, game, and network theory and the rapidly growing research field of complex systems and their simulations. The topics range from simple coupled dynamics, via the principles of chaos theory, the theory of Cellular automata, game and network theory to methods of artificial learning, the theory of artificial neural networks and the design of self-referentiality and autopoiesis in complex systems. At the same time, Füllsack tries to mediate between the Anglo-American research tradition (The Santa Fe Institute, Bob Axelrod in Michigan, Joshua Epstein of the Brookings Institution, and Steven Strogatz and Duncan Watts) with the German systems theoretic debates in the wake of Niklas Luhmann. VS-Verlag, Wiesbaden, 2011. ISBN 978-3-531-92765-7. 335 pages.
Are Computers Digital? Should Constructivists Care?
Stefano Franchi
Texas A&M University, USA
stefano.franchi/at/gmail.com

Open Peer Commentaries on Manfred Füllsack’s “Constructivism and Computation”

Are Computers Digital? Should Constructivists Care?
Stefano Franchi
Texas A&M University, USA
stefano.franchi/at/gmail.com

> Upshot • While I do agree with Füllsack’s positive assessment of the use of computer simulations in advancing constructivism’s program, I am less convinced by the alleged opposition between computers and constructivism he builds up. In my opinion, his depiction of computers and computation is inaccurate in several respects. As a result, the alleged incompatibility with constructivism Füllsack objects to disappears, and his whole essay could then be construed as a classic straw man fallacy. This is unfortunate, because it could lead to an unwarranted dismissal of his positive contribution.

« 1 » Manfred Füllsack argues that the use of computer models can help constructivism’s scientific and philosophical program. An extended review of several works in fields as diverse as contemporary robotics, the epistemology of artificial life, and computer simulations of extended versions of classic game-theoretic quandaries (such as the prisoner’s dilemma) allows Füllsack to conclude that:

What I tried to show in this paper is that constructivism and computation do not contradict but, on the contrary, are able to enrich each other productively. Computer-based models such as the one on downward causation cannot prove constructivism, but they can open up respective constructions to detailed analysis. They allow for a step-by-step investigation of interactions that remained widely impenetrable prior to the age of digital computers.** ($§5$, my emphases)

« 2 » I am not convinced by the opposition between constructivism and computation that provides the starting point of Füllsack’s argument. As I will argue below, the alleged incompatibility relies on an inaccurate or altogether missing definition of the term “computers.” However, Füllsack’s discussions of computer-based works provide positive evidence for the usefulness of simulation in constructivism that is logically independent of the flawed computer-constructivism opposition that frames his work. Separating the positive claims from the overall logical framework would only strengthen them.

« 3 » In the opening sections of the essay ($§1$–$2$), Füllsack constructs the argument he wants to disprove as follows:

1 Constructivism holds that there is no unobserved reality or no unmoved mover.
2 Computers are deterministic and discretely operating machines.
3 Computers are subject to the grounding problem: they require an externally given “first,” usually defined by a program who is not part of the computation.
4 Therefore, computers are ill-suited to constructivism, insofar as they reject its most basic tenets and cannot model truly self-referential, end-directed systems.

« 4 » I think premises (2) and (3) are either vague enough to defy a proper assessment or they become outright false or trivially true when their meaning is properly narrowed. As a result, the conclusion, (4), does not carry through, and the whole argument – which the author sets out to refute in the remainder of the essay – turns into a straw man fallacy.

« 5 » I will start with statement (2). What is Füllsack referring to, when he says that “computers” are “deterministic” and “discretely operating” machines? Is he pointing to the machine itself – the hardware (the “bare metal” as people sometimes say)? Or is he talking about the abstract logical model – the Universal Turing machine – that a particular physical computing device imperfectly approximates? Perhaps by “computers” he means the conjunction of the two: a specifically individuated physical device plus the abstract logical model. Or perhaps Füllsack intends “computers” to refer to particular computing architectures that are built on top of the physical device plus abstract logical models, such as von Neumann architectures, or even higher-level structures such as cognitive architectures built on top of von Neumann architecture built on top of Turing machines built on top of physical hardware. Which one of these many possible referents for “computers” is “deterministic” and “discretely operating”? The hardware, it seems to me, is as deterministic and discrete as any other spatio-temporally located chunk of the world. A computational device’s hardware is no different, in principle, from a combustion engine – or from a cup of coffee, for that matter. Is the latter discrete and deterministic? Perhaps it is it and perhaps not – it all depends on whether we consider the world as ultimately made up of deterministic acting particles (as a sort of Laplacean universe, so to speak), or of strings, or quarks, or fields, and so on. It also depends, of course, on whether we admit that the sentence “The world is ultimately made up of x” is philosophically meaningful. Constructivism
would deny it, as Füllsack correctly stresses. Be it as it may, computers, in this first mentioning of the word, cannot be properly differentiated from the rest of the world we interact with. Hence, statement (2) cannot be made to be true — it remains vague and undetermined.

- **6** Someone could object that my equating CPUs and RAM to coffee cups misses the point entirely: computer hardware may not be really digital, the objection would go, but it is designed to behave as if it were digital. The physical counterpart of a bit of information stored in RAM is really the electric charge held by a capacitor plus a resistor. That charge varies continuously, not discretely, even though the introduction of threshold levels allows the “user” (i.e., the operating system) to treat the continuously changing capacitor plus resistor combination as a binary digit. However, the very fact that electrical engineers do indeed spend a good deal of their time trying to overcome the non-discrete nature of the material they deal with when they design computer hardware proves my point, it seems to me. Computer hardware is not digital, even though, under the appropriate assumptions, it may approximate digital behavior to a certain degree. The discussion of a computer’s intrinsic discreteness must then move from the hardware to the software that runs on it and, more specifically, to the logical models underlying the software. More precisely, the discussion of discreteness should move from the ontology of computers to an ontology of computation. And this is not an easy task at all.

- **7** I will just mention some of the issues that immediately arise. Turing machines are both deterministic and discrete by design, and for obvious reasons. However, we could say that (2) is true if we replace “computers” with “Universal Turing machines.” But the argument’s conclusion becomes much less interesting, as (4) would have to be rewritten as “Universal Turing machines are not self-referential end-directed systems.” This strikes me as probably true, but I doubt many people would be surprised by it: Turing designed his abstract machine not to be self-referential and end-directed. Even this restated conclusion, though, hides two problematic assumptions.

- **8** First, it implicitly equates computation (in the abstract logical sense) with Turing-computation by implicitly defining a computable function as any function whose value could be computed by a properly set up Turing machine. This assumption has come under renewed scrutiny since Jack Copeland and Diane Proudfoot’s (2004) discussion of the various computational models Turing himself discussed in his later work. While non-Turing computers (or “hypercomputing machines,” as they are sometimes called) may be still far off in the horizon, it is seems safe to say that we should exercise extreme caution when identifying computation in general (and, as a consequence, “computers”) with Turing-computability.

- **9** Second, the fact that Turing-machines are not self-referential, end-directed systems does not necessarily mean that they could not be used to model some interesting subset of the latter’s behavior, thereby helping us to understand them better. Let me use a common example to illustrate the point. Even though a mathematical model of a storm may not get you wet and it is therefore not a storm in a very relevant sense of the word “to be,” we can still learn a lot about real storms by running the model. Thus, even if it were clear that computations are discrete and deterministic (and we should at least doubt the truth of this statement as soon as we move away from abstract Turing machines into real computations), we could not exclude that sufficiently complex Turing machines could serve as suitable models of (some aspects of) possibly non-deterministic, self-referential, continuous, end-directed processes. Indeed, Füllsack’s aim in the remainder of the essay is devoted precisely to providing positive evidence for the truth of this claim. The other possible readings of statement (2) I listed above (as a combination of hardware plus abstract models, at various levels of logical complexity) run into problems analogous to those just discussed and I will not go into further details.

- **10** Let me now look at statement (3): “Computers are subject to the grounding problem.” Here again, the main problem lies with the reference of “computers.” In his 1990 essay, Stevan Harnad offered the symbol grounding problem as the ultimate obstacle against any psychological theory (such as classic GOFAIs) that reduces cognition to symbol manipulation. “How is symbol meaning to be grounded in something other than just more meaningless symbols?” he asks. Symbols must be grounded in something other than symbols: they must receive their meanings from the subject’s direct experience of the world. Claiming that cognition amounts to symbol manipulation is tantamount to claiming that we could learn “Chinese as a first language and the only source of information you had was a Chinese/Chinese dictionary! [...] This is the symbol grounding problem!” (Harnad 1990: 339–340).

- **11** Harnad’s critique had a very specific target: the theory (still very influential in 1990) that identified cognition with symbol manipulation, and the mind as a (physical) symbol system. Hence, accepting Harnad’s argument would only lead us to a limited and hardly surprising conclusion. Namely, that classic cognitivist theories of the mind such as those originally proposed by Herbert Simon and Allen Newell and later made popular by Jerry Fodor are incompatible with constructivism. Harnad’s critique tells us very little about computers in general.

- **12** One could perhaps try to extend the symbol grounding problem to computers in general by arguing that the only possible way to ground symbols’ meaning is by having a body causally connected to the world. Symbol grounding, in other words, requires embodiment, as proponents of embodied cognition paradigms have repeatedly argued in the last twenty years. Computers are not embodied, hence they will never be grounded. But is this last statement true? First, the considerations I offered above apply in this case as well. (Universal) Turing machines are certainly not embodied, but computers are not Turing machines. A physical instantiated approximation of a Universal Turing machine does not live in the same abstract
space as set theoretic statements and logical formulas.

« 13 » Still, being a part of the spatio-temporal continuum is not equivalent to having a “body,” one could reply, perhaps leveraging the philosophical definitions of a living body offered (among others) by Maurice Merleau-Ponty or by Hans Jonas. Lizards and humans have bodies. Plants and bacteria may have them. Rocks certainly do not. The objection carries a certain weight, but its import about the status of “computers” is less conclusive than it may seem. What if we could give computers a body? What if we could allow computers to interact autonomously with the world on the basis of their internal needs? Would those computers still be suffering from the symbol grounding problem? My question is rhetorical, of course: embodied computers are called robots. That we still do not know how to build fully autonomous robots is not important. As long as we do not have conclusive evidence that such creatures cannot be built, the assertion that “computers suffer from the symbol grounding problem,” cannot, in my opinion, be taken as true.

« 14 » Moreover, the same epistemological issue about computers being (models of) a cognitive agent I mentioned above reappears. One could grant that (future) autonomous embodied robots may be self-referential, end-oriented beings and still deny that ordinary devices such as the desktop computer with which I am writing this paper will ever be one. Even if this last statement were unconditionally true, it would still be the case that I can use a disembodied, non self-referential, non-end-directed desktop to simulate (i.e., to model) certain aspects of autonomous robots’ behavior. In fact, this is precisely how a large part of everyday work in robotics is routinely carried out (including precisely how a large part of everyday work autonomous robots’ behavior. In fact, this is

« 16 » Yet my critique does not touch the substance of Füllsack’s work, only its logical presentation. Determining if “computers” are incompatible or not with constructivism’s insights is not important. A constructivist would rather claim that what is important is to use them to further our understanding of ourselves. In this respect, I think Füllsack’s work succeeds. The computer simulations he discusses shed light, as he claims, on several questions about the “constituencies of life, the functioning of cognition and the emergence of mentality” (§2). His claim would only be strengthened by removing its dubious link to computers’ “unconstructivist” nature.

Stefano Franchi is Associate Research Professor in the Department of Hispanic Studies, Texas A&M University. He works on twentieth century European philosophy and the history of artificial intelligence and cognitive science. He is the editor (with Güven Güzeldere) of Mechanical Bodies, Computational Minds (MIT Press 2005) and (with Francesco Bianchini) of The Search for a Theory of Cognition (Rodopi 2011). He is completing a monograph on contemporary alternatives to Hegel’s theory of historical development, provisionally titled Play and Passivity.

Weak and Strong Constructivist Foundations

Marco C. Bettoni
Swiss Distance University of Applied Sciences (FFHS)
marco.bettoni/at/weknow.ch

> Upshot - Füllsack’s article offers many interesting ideas but falls short of elucidating the relationship between constructivism and computation. It could profit by taking into consideration stronger constructivist foundations such as the distinction between machine and organism, the relationship between reality and the observer, and Ceccato’s theory of attention.

“it is the coherence of experiences with other experiences that constitutes the foundation of all explanation”
Humberto Maturana in Maturana & Poerksen (2004: 42)

« 1 » Manfred Füllsack introduces and outlines some interesting and useful ideas. For example in §18, the Bayesian brain approach of contemporary robotics is said to confirm Piaget’s consideration of sensorimotor interaction as being circular and recursive; further, in §32, Mark Bedau’s definition of “emergence” in terms of computer-based simulation according to which emergent properties are not reducible other than by way of computation is outlined. There is also the simulation of “downward causation,” with Joshua Epstein’s demography prisoner’s dilemma (§39ff.), which merits further attention. This is especially so where the author discusses a “cooperation model” that explains aspects of the “coevolution of system and environment,” for example “emergence” and “immergence.” However, Füllsack claims that “constructivism and computation do not contradict” (§55) does not appear to be sufficiently justified. Could this be due to the weak constructivist foundations on which the paper is based? For this reason, my comments will focus on an attempt at strengthening these foundations by pointing at three essential issues: a distinction between machine and organism, the relationship between reality and the observer and Silvio Ceccato’s theory of attention.

« 2 » In §8, the author claims that a “close connection” between constructivism and computer-based modelling becomes clear when simulation illustrates how “realities sui generis (a.k.a. systems) can emerge.” In my view, this illustration is useful but not relevant nor essential since a sophisticated machine could do the same. What we need instead is a convincing, viable distinction between an organism (living organisation) and a machine (dead organisation). A good source of inspiration for such a distinction was provided more than 200 years ago by Immanuel Kant (1981), who discusses the issue in detail and finally suggests that “An organized natural product is one in which

http://www.univie.ac.at/constructivism/journal/9/1/007fuellsack
everything is an end and reciprocally also a means" (A292, my translation). Obviously, this is not the case in a machine, since, as Kant himself suggests in one of his rare examples:

\textbf{3} In Kant's distinction between machine and organism, I see a prelude to Humberto Maturana & Francisco Varela's concept of autopoiesis (Maturana & Varela 1980: 79). It was this concept that in the following years enabled Maturana to provide further developments (such as the criterion of validation of scientific explanations, the distinction of two explanatory paths of objectivity, the operations of distinction of the social and of the ethical, etc.) concerning the relation between reality and the observer (see next). In my view, they constitute some of the most essential, strong foundations of constructivism.

\textbf{4} In an earlier work of mine, for instance, where I characterized the cognitive system as an organic system, I was inspired by autopoiesis to specify it further as follows (Bettoni 2005: 18; Bettoni & Eggs 2010: 134). What we construct in the act of knowing (doing) can be fed back to the cognitive functions, and this feedback is such that the knowledge fed back becomes a component of the cognitive functions that produced it: thus knowledge is not just a result! Thanks to the feedback, the new construct becomes an integrated system element and builds on, extends, enhances the potential, the cognitive means, the being of a person. In the language of cybernetics, we would talk of the operation (doing), operand (knowledge) and operator (being); the special thing here is that the feedback becomes a function that expands the mechanism from which it originated. And so the whole system grows (dotted lines in Figure 1), i.e., when the system is active, it also grows in its capabilities. Doing generates being: this is how I interpret “autopoietic.”

\textbf{Maturana about reality and the observer}

\textbf{5} In a well-known article dating from 1988, Maturana states that “the most central question that humanity faces today is the question of reality” (Maturana 1988: 25). However this does not appear to be the concern of the target article, which instead focuses throughout the text on circularity, and at the end (§55), characterizes the constructivist conception of knowledge as “circular”; but if circularity is so important, what is its relation to the question of reality?

\textbf{6} In the end, realism is not expressively refuted (§55), leaving the choice between realism and constructivism “a matter of taste.” This is an attitude that shines through the whole text but in my view prevents the authors from making progress in solving the problem of reality in a more viable way. In the constructivist thesis that “reality is the construction of an observer” and in the realist thesis that “reality exists independently” (§4), the term “reality” is merely a homonym that means two different things; hence it is confusing to mention them as if they would mean the same thing. For me as a constructivist, the reality that I construct is not a physical reality but a conceptual one. For a realist, on the contrary, the reality that he sees as existing independently is actually a combination of the two: physical reality as the “reference” and conceptual reality as its “copy” in his head. In constructivism, we need to disentangle these two things, and here is where Maturana’s reflections about autopoiesis and cognition become essential. Consider, for example, his differentiation between two fundamental ways in which we, as humans, can understand explanations; he distinguishes two mutually exclusive explanatory paths: the path of “objectivity without parentheses” and the path of “objectivity with parentheses” (Maturana 1988: 28ff).

\textbf{7} On the explanatory path of objectivity without parentheses, the observer assumes either implicitly or explicitly that he is capable of making statements about the logic of things, as if the logic he accords to them would exist independently of him. He does not ask himself: “How can I say that the logic of this thing exists independently of me?” If someone makes the implicit assumption that he can reference things, as if the logic he accords to them exists independently of him, then he is also effectively stating that the explanations he applies can ultimately be validated by the things themselves, independently of him. This explanatory path therefore contains the implicit and unaware assumption that an individual can reference a logic that exists independently of him and that validates what he says. And what could that be? It is a logic of reality (the logic of being, the essence of things, etc.) or, in other words, a universal truth. It is universal because it exists independently of us. It is valid for everything because it is independent from everything.

\textbf{8} On the explanatory path of objectivity with parentheses, the observer notes something different, something very interesting: that his explanations are validated by his actions. The logic of his experience is explained by the logic of his other experiences and not by a reference to a logic that is independent of us. The observer sees himself as a source of validation for his own statements. This is the essence of the essence of constructivism! According to Maturana,
parentheses around “objectivity” represent awareness – that special awareness that we, in order to validate our explanations, are unable to refer to or reference anything (neither a thing nor a logic) that is independent of us. Plus an awareness of the fact that our explanations are validated by coherence in the logic of our experiences: “In fact, scientific explanations do not explain an independent world, they explain the experience of the observer” (Maturana 1988: 38).

- « 9 » By being aware that we explain the logic of our experience through the logic of our experiences, we notice that there are numerous domains of explanation because each domain of coherent experiences represents a domain of explanation: in this domain, we can draw on experiences within it in order to explain other experiences from it. And since each of these domains of explanation is experienced as a domain of objects or as an area of reality, this explanatory path gives rise to numerous realities, even if the physical reality remains one. This is because they are “realities for me” and not absolute realities. But how could we make this step and reach the awareness mentioned by Maturana?

**Ceccato’s “Theory of Attention”**

- « 10 » Something that could help us become more aware of our construction of reality is, for example, Ceccato’s “Theory of Attention,” inspired by Percy Bridgman’s operationalism. 8 His “operational idea” was the starting point for a development in the theory that led to approaching concepts as operations (operational analysis of concepts), one of the core thoughts behind radical constructivism. Physicist Bridgman discovered that the problem of simultaneity in Einstein’s theory of relativity can be elegantly resolved by defining the concept of simultaneity by means of operations. This approach from physics also fitted perfectly with Ceccato’s idea that we construct concepts through mental operations (Ceccato 1947): Ceccato started to devise an analysis of concepts that could identify the mental operations needed to generate a concept.

- « 11 » According to Ceccato’s theory, the actual fundamental conceptual operations are not physical actions, as in physics, but rather “moments” or “states” of attention. Attention is usually presented in psychology or common sense as a kind of spotlight that illuminates something. In the case of Ceccato, attention is a much more comprehensive function (an attentional system or organ), which has a special “constitutive” and also “regulative” impact rather than being purely image-based: the operations of attention determine the object as far as its logic is concerned, and not the other way around. Attention is the mechanism by means of which we create our constructs, our reality; the How (attentional operations) determines the What (our reality), as far as its logic is concerned.

- « 12 » The functionality of this proposed attention organ, which I call the “categoriser,” is derived from a pulsating fundamental notion: the categoriser produces an uninterrupted and even rhythm of (“conscious”) moments or states, rather like our breathing and circulation. Experiments in neuroscience have indirectly confirmed this approach on several occasions (Harter 1967; Lehmann et al. 1998). Dietrich Lehmann and his research group in Zürich, for example, have devised experimental results that suggest that “the seemingly continuous stream of consciousness consists of separable building blocks.” These attentional blocks or moments serve as building blocks in the construction of more complex units (Bettoni 1989: 13). Through mental operations (so-called “categorising”), we construct these units using the categoriser, combining the moments with one another (free moments) and with the functioning of other organs, e.g., sensory impulses i.e., the eyes, ears, etc. (focalised moments). Ceccato gave the name of “mental categories” to these connections between free moments of attention, in honour of Kant (Ceccato & Zonta 1980: 53). Examples include pure terms such as “something,” “object,” “and,” “or,” “with,” “singular,” “start,” “end,” “element,” “point,” “true” and “energy,” etc. Examples of connections from focalised moments are “hard,” “water,” “horse,” “melon,” “paper,” “pencil,” “cat,” “guitar” and “sun,” etc.

- « 13 » In §28 Füllsack mentions the focus of European epistemology on substances and objects rather than on “processes that bring them about.” This seems to be a good place to link with Ceccato’s theory of attentional operations; but the focus must be on a functional level, not on the recursive processes mentioned in the article, because these belong merely to the implementation level.

**The Construction of the Environment**

Bernd Porr
University of Glasgow, UK
bernd.porr[at]glasgow.ac.uk

> Upshot • The environment is not slowly constructed by the agent but is an integral part of being an agent because both, agent and environment, are part of a closed loop system. By identifying the perturbations impacting on the loops, with the help of second-order cybernetics, the agent can identify them as its environment.

- « 1 » My main criticism of this article is the establishment of the environment (see §§ 24 and 25). In the classical constructivist tradition there is, of course, no “real” environment and it needs to be constructed. Through the interaction with “something,” we construct the environment. In my initial review, I dared to say “real environment” and the author rightfully reminded me of the dispensable nature of the “real world” in constructivism. I am now going to describe
in greater detail my definition of the “environment” and why the target paper is not radical enough when defining it.

“2” What is environment? Clearly it needs to be different to the agent (Alrøe & Nøe 2012)? How to avoid being branded a naive realist when talking about this “something” called environment that seems to be somewhere? Like AI, or biologically-inspired robotics too, constructivist theory has “a grounding problem” in that it bases itself on – constructivism. So, what is one allowed to say before one faces getting thrown into the unmarked space of the non-constructivist limbo?

“3” From my point of view, there are two main aspects:

1 A computer needs to run in a closed loop, for example a robot’s actions need to feed back to its sensors, and so on. Perturbations arise in the electronics, or in the environment (for example, obstacles, food, attacks, other agents, etc.)

2 The robot needs a mechanism to develop new loops, for example for predicting perturbations so that it is able to observe with these loops the previously created loops. This can easily be done by closed loop learning (Porr & Wörgötter 2006), which controls desired states and tries to predict perturbations in other loops.

3 Once the prediction of perturbations has been established, the robot can identify them as environment, be they food or another agent, in the sense of Luhmann’s double contingency.

“5” Therefore, the problem in this article is that environment is not clearly defined and is just introduced as “something” that needs to be constructed but outside of the agent. It appears that the agent is first created as an entity and then slowly “constructs” its environment while interacting with it. The approach needs to be more radical. The agent and the environment are one entity forming a closed loop from the outset. The environment is not a fuzzy entity at the beginning that then gains shape, but is rather as solid as the agent from the beginning. Coming back to von Foerster: the central heating works from the outset; it has a construction of the weather in terms of a maximum boiler power, which in turn represents the requisite variety of the loop dictated by the environment. It could be improved by adding more control loops that improve the first one through learning. The same applies to an animal that can successfully escape from a predator. Here, its maximum muscle strength and its strategies to run and then hide are its constructions of the environment and can be improved by observing its environment while it is escaping or on reflection. On the level of social systems, one could, for example, think of the financial system, where a sudden loss in profit requires a company to adjust where its construction of its environment takes into account the factors that impact on their internal operations, factors such as changes in customer behavior or the rise of a competitor. Again, the environment was part of the loop from the outset because a company can only run its self-referential operations when its loops are working from the very beginning. Coming back to the digital computer: it can act both as the agent and as the environment as long as we obey the rules set out above and we accept that the loops we create model the loops in animals that have been created by evolution. However, these could be created with the help of genetic algorithms that select agents that have working loops, for example the agents that can escape a predator or can find food. In the end, this is how Humberto Maturana founded his constructivist approach, namely requiring stable, self-referential systems (i.e., with attractors) that are able to maintain their self-referentiality throughout their life. For example proteins produce proteins and so on in a self referential cycle and they do not diffuse into the ocean, which essentially means that they should not die.

“6” In summary, I find that the article is not radical enough in how environment is introduced. Environment and agent represent one unit forming a closed loop. The agent needs to tease out what is environment and what is agent by observing perturbations acting on the closed loops to identify the environment, which is essentially the source of perturbations.

Bernd Porr has degrees in physics and journalism and a Ph.D in computational neuroscience from the University of Stirling. Since 2004, he has been a lecturer in electronics and electrical engineering at the University of Glasgow. His research interests range from neurophysiology, through biologically-inspired robotics, to social systems.

Accepted: 23 October 2013

Received: 11 October 2013
Author’s Response: Constructivism as Possibility?  Manfred Füllsack

> Upshot • Does constructivism need to assert its validity or is it more appropriate to assume its possibility, discuss its consequences and try to deliver arguments that show it is a viable epistemology?

« 1 » As all three commentaries address issues that, from my point of view, were not intended to be centrally addressed in my paper, I conclude that my way of expression was deficient at least over parts of the text. So let me take the opportunity of this response to once more to indicate what my original intention was.

« 2 » I agree with Stefano Franchi’s criticism that the theoretical conception of computation in my paper is vague. Indeed, I have to admit, I do not have any elaborate theory of computation in the way that Franchi demands, even though my initial statements about Stevan Harnad’s grounding problem (§1), Tipler’s universe simulations (§5), or the Zuse-Fredkin-Thesis (§8) might convey this impression. What Franchi (§3 and §5) calls my “statement 2” about computers being deterministic and discretely operating machines simply builds on the somehow profane experience of deploying the technique of computer-based modeling to simulate the interaction of complex dynamics. In this respect, it builds on the possibilities of the machine itself rather than on any elaborate logical conception of Universal Turing Machines, Von-Neumann architectures or other abstract conceptions. In other words, my use of the term “computer” simply stems from the experience of modeling, that is, for example, of having to think in terms of difference instead of differential equations (respectively, their numerical instances) when deploying computers to model dynamics that are perceived as continuous. This is why I started my paper with a proposition about a perceived contradiction between computation and constructivism. The exact wording is: “modern-day digital computers seem to contradict the foundational assumption of constructivism” (§1, emphasis added). For me, the reference to semblance matters. In no way did I intend to purport any ontological “opposition between constructivism and computation” as Franchi assumes. For this – I agree – an elaborate theory of computation would be needed. But this was not in the focus of my paper. What I meant to say was just that if someone (like me) perceives computers as deterministic, discretely operating machines with a grounding problem, then this must not necessarily pose a problem to considering the output of computer-based models as contradictory to constructivism.

« 3 » The same goes for the remarks on the symbol grounding problem (§§10–15). I did not state that “computers suffer from the symbol grounding problem,” as stated in Franchi §13. The exact wording is that I consider them subjected to the symbol grounding problem. While this formulation is not as clear as would be needed to express the difference between “semblance” and “being,” it is still a difference that makes a difference in terms of constructivism. The symbol grounding problem can be perceived as indicating a problem that seems to contradict basic assumptions of constructivism – that is what I wanted to express.

« 4 » I cannot exclude the possibility that with this cutback, “the argument’s conclusion turns out to be trivially true,” as Franchi (§15) argues. The aim of my paper however, was not to state that “computers contradict constructivism.” Instead it was to show that computer-based modeling and simulation can support the case for constructivism even though the bare technical endeavor of writing a computer simulation encounters aspects such as discreteness, determinism and grounding problems that might be seen as contradicting constructivism.

« 5 » Somehow, symmetrically to the proposition of Franchi about my use of a deficient conception of computation, Marco Bettioni in his opening statement rejects my concept of constructivism as too weak to support the claim that constructivism and computation do not contradict. He expresses the need for a “convincing, viable distinction between an organism (living organisation) and a machine (dead organisation)” (§2) and suggests referring to Immanuel Kant as a precursor of Humberto Maturana & Francisco Varela and their autopoiesis conception. This then gives rise to sketching a model of autopoietic knowledge, as suggested by Bettioni himself. As interesting as this sketch might be, I do not quite understand how it and the respective thread of reasoning could strengthen my concept of constructivism. Unfortunately, the same seems to be true for the suggestion to draw on Ceccato’s “Theory of Attention” in order “to become more aware of the construction of reality”. My lack of comprehension might stem from the fact that I do not feel unaware of the possibility that reality is a construction. Neither (but this might be subjective) do I consider my text to be prone to such unawareness (see §1, §4, and §10, where I refer to this central aspect of constructivism). The reason why I remain cautious on statements about reality being a construction (and therefore, in this respect, maintain a rather subjective mood throughout my text) simply has to do with the fact that in my opinion ontological statements do not go together well with constructivism. Being a constructivist, I cannot exclude the possibility that the insight that reality is a construction is a construction itself. That is the relativistic intricacy I mention in §4, and the reason why I prefer to keep to the proposition that choosing between realism and constructivism remains “a matter of taste.”

« 6 » Finally, Bernd Porr in his commentary expresses dissatisfaction with the lack of radicality of how I deal with the issue of environment in my paper. From my point of view, this objection is subject to a similar misreading as it reverberates in the commentaries of Franchi and Bettioni. I did not intend at all to elaborate on the ontology of environment in my paper. From my point of view, this objection is subject to a similar misreading as it reverberates in the commentaries of Franchi and Bettioni. I did not intend at all to elaborate on the ontology of environment in my paper. As I said before, I would prefer to withhold any final verdict about ontological issues. I think that even with a clear inclination towards constructivist epistemology, it is not necessary – and it is no good constructivistly standing – to express any such final verdicts in science. All I wanted to show in my paper is that if one decides to take a constructivist’s stance and wants to consider the environment as co-emerging or co-evolving with the system (or the agent, respectively), then computer-based modeling (or computation, respectively) does not pose an insurmountable obstacle to doing so. As my version of the demographic prisoner’s dilemma should demonstrate, it is not only possible but even
Combined References


