Becoming Digital: Reconciling Theories of Digital Representation and Embodiment

Harry Halpin

1 Introduction

One of the defining characteristics of information in actually-existing computational mechanisms ranging from the World Wide Web to word-processors is that they deal in information that is - or at least seems to be - robustly digital, bits and bytes. Yet shockingly, there is no clear notion of what 'being' digital consists of, even though a working notion of digitality is necessary to understand computers, if not human intelligence. This is not to say that 'digitality' is not understood in a practical or engineering sense, for assuredly we build digital systems. While engineers can implement digitality, and ordinary people 'know it when they see it,' there is no rigorous philosophical definition of digitality. So a whole host of questions are left unanswered when human intuitions over digitality vary, which can easily happen outside of a practical engineering context. For example, are concepts digital? Can non-human artifacts be digital? Is digitality subjective or objective? [22]. These kinds of questions can not be answered rigorously because philosophy has in general ignored inspecting the intuitions behind digitality, so our first task should be to create a philosophical definition of digitality.

Furthermore, much of the power of computation comes not only from digitality, but from the ability of computers to 'represent' things. Again, the situation is similar to digitality: namely, that almost anyone can 'spot' a representation when they see one, such as a picture of the Eiffel Tower or the words 'Eiffel Tower.' Unlike digitality, representations have been a core topic of philosophical investigation in cognitive science [6]. However, over the last twenty years a movement against digital representations has been gaining momentum in the field of artificial intelligence (AI). This movement usually goes under the slogan of 'embodiment,' as many researchers wanted to move the focus of AI to more biologically realistic work around dynamical systems and neural networks [3]. While once a minority

W3C/MIT
32 Vassar Street Room 32-G515 Cambridge, MA 02139 USA, e-mail: hhalpin@w3.org

within AI, at this point anti-representationalists are the clear majority. Their philosophical lineage can primarily be traced to Hubert Dreyfus's Heideggerian analysis of intelligence, which rejects the role of representations in intelligence altogether [9]. Another more subterranean anti-representationalist influence is the theory of autopoiesis of Maturana and Varela [21]. These strands of anti-representationalist philosophy have rejected the possibility of computationally-implemented artificial intelligence on a priori metaphysical grounds. However, more empirically-inclined philosophers such as Clark [3] and Wheeler [29] have revived the philosophy of artificial intelligence with many of the insights of embodiment while still holding out for artificial intelligence as an engineering possibility. Influenced by this philosophical stance, most researchers have adopted an anti-representationalist stance in their practical work towards building artificial intelligence, such as the well-known work of Rodney Brooks in robotics [2]. Yet, surprisingly, very little of this work has come to fruition: Brooks is well-known for having simulated animals, but his project to simulate actual human-level intelligence seems to have stalled. Not to mention that there is a movement to incorporate the environment into the task of both philosophical and engineering investigations of intelligence, as exemplified by the work around the Extended Mind Hypothesis by Clark and Chalmers [5]. However, these researchers have yet to come to grasps with the fact that this wider environment would definitely include computers, the Web, and other rather intuitively information-carrying digital representations. Previously, almost all work in the philosophy of AI has focused on debates over the possible existence of representations that are assumed to be implemented neurally. We can remain agnostic on this question while at least accepting that representations do exist external to the neural system. Thus, our second task should be to define a definition of representation that is *independent* of whether a given representation is internal or external to the human body as conventionally defined by the barrier of the skin. Lastly, our explanations of representations and digitality must be purely causal so not incompatible with the strict materialism that is necessary for a scientific understanding of embodied and embedded intelligence.

2 Preliminaries

On the surface a term like 'representation' seems to be what Brian Cantwell Smith calls "physically spooky," since a representation can refer to something with which it is not in physical contact [27]. This spookiness is a consequence of a violation of *common-sense* physics, since representations allow us to have some sort of what appears to be a non-physical relationship with things that are far away in time and space. This relationship of 'aboutness' is often called *reference* or *intentionality* and is considered to be the defining characteristic of representations. While it would be premature to define 'representation,' a few examples will illustrate its usage: someone can think about the Eiffel Tower in Paris without being in Paris, or even having ever set foot in France; a human can imagine what the Eiffel Tower would look like

if it were painted blue, and one can even think of a situation where the Eiffel Tower wasn't called the Eiffel Tower. Furthermore, a human can dream about the Eiffel Tower, make a plan to visit it, all while being distant from the Eiffel Tower. Intentionality also works temporally as well as distally, for one can talk about someone who is no longer living such as Gustave Eiffel. Despite appearances, intentionality is not epiphenomenal, for intentionality has real effects on the behavior of agents. Specifically, one can remember what one had for dinner yesterday, and this may impact on what one wants for dinner today, and one can book a plane ticket to visit the Eiffel Tower after making a plan to visit it.

Can we get to the heart of this mystery of representation without recourse to some kind of dualism? The trick would be to define what precisely our common-sense notion of representation is, and to do this requires some terminological ground work while avoiding delving into amateur quantum physics. The terminology here is supposed to reconstruct rather carefully some common-sense demarcations in an uncontroversial yet broad manner. To pin the supposed 'spookiness' of reference down, we will introduce a few terms. A ***process*** - or 'thing' - is a general-purpose term used to denote events, objects, and proto-objects in a "patch of metaphysical flux," where a thing can be defined by having some regularity in time and space that can distinguish it from other possible things [27]. A ***regularity*** is a lack of difference in time and space at a given level of abstraction. There are generally two kinds of separation possible in processes in a relativistically invariant theory, a physical theory that obeys the rules of special relativity so that the theory looks the same for any constant velocity observer, as processes may be separated in time or space. Things that are separated by time and space are ***non-local*** (disconnected) while those things that are not separated by time and space are ***local*** (connected). While a discussion about counterfactuals and causation is far beyond our scope, we will rely on the common-sense intuition that if one process is connected with another thing and a change in the former thing is followed by a change in the latter thing, that former process may have caused the change in the latter process. Anything that appears to violate these common-sense intuitions about physics and causation is ***spooky***, while anything that does not is ***non-spooky***. A property of the distal is that it is beyond effective reach; as Smith puts it, "distance is where no action is at" [27].

3 Information, Encoding, and Content

In order to define digitality and representation, we will have to reformulate the notion of information, building on Shannon's information theory [25]. To rephrase as best as we can the mathematics of Shannon in natural language, ***information*** is *whatever regularities held in common between two processes*, a ***source*** and a ***receiver*** [25]. To have something in common means to share the same regularities, e.g. parcels of time and space that cannot be distinguished at a given level of abstraction. This definition correlates with information being the inverse of the amount of 'noise' or randomness in a system, and the amount of information being equivalent

to a reduction in uncertainty. This preservation or failure to preserve information can be thought of the as sending of a message between the source and the receiver over a channel. Whether or not the information is preserved over time or space is due to the properties of a physical substrate known as the ***channel***.

Shannon's theory deals with finding the optimal encoding and size of channel so that the message can be guaranteed to get from the sender to the receiver [25]. Yet, what is encoding? Goodman defines what we would call an encoding as a series of marks, where a ***mark*** is a physical characteristic, such as the marks on paper one can use to discern alphabetic characters to ranges of voltage that can be thought of as bits [12]. To be reliable in conveying information, an encoding should be physically "differentiable" and thus maintain what Goodman calls "character indifference" so that (at least within some context) each character (characteristic) can not be mistaken for another character. So, an ***encoding*** is a set of precise regularities that can be realized by the message.

There is more to information than encoding. Shannon's theory does not explain the notion of information fully, since giving someone the number of bits that a message contains does not tell the receiver *what* information is encoded. Shannon himself explicitly states, "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem" [25]. Many intuitions about the notion of information have to deal with not only how the information is encoded or how to encode it, but what a particular message is about, the *content* of an information-bearing message. 'Content' is a term we adopt from Israel and Perry [19], as opposed to the more confusing term 'semantic information' as employed by Floridi [10]. Floridi rejects traditional Shannon information theory in favor of constructing his own idiosyncratic theory of 'semantic' information, but his rejection is based on a common misunderstanding of Shannon's information theory as merely a theory of communication between a source and a receiver. However, the receiver and sender can exist over time rather than space, and so be the same physical object. For example, information (such as my eye color) is preserved (and can even be thought of as a message!) between myself at five-years old and myself at thirty-three years old. Information is not about communication, but about the preservation and determination of structure, which is necessary both for digitality and representation to work. Not to mention that logic-based AI has essentially been superseded by machine-learning in artificial intelligence, and machine-learning is firmly defined in terms of Shannon information theory.

Structure is needed to convey content, but what is content? While the notion of an informational content is hard to pin down, it is easy to illustrate. Let's imagine the case where we are trying to deliver the message that Ralph, a single employee at a company that has eight employees, won a trip to Paris. Just determining that Ralph won a free trip to Paris requires at least a three bit encoding and does not tell us which person in particular won the lottery. Shannon's theory only measures how many bits are needed to tell us precisely who won. After all, the false message that

tells us wrongly won a trip to Paris is also three bits. Yet content is not independent of the encoding, for content is conveyed by virtue of a particular encoding and a particular encoding imposes constraints on what content can be sent [25]. Let's imagine that Daniel is using a code of bits specially designed for this problem, rather than natural language, to tell us who won the free plane ticket to Paris. The content of the encoding `001` could be Ralph while the content of the encoding `010` could be another employee, Sandro. If there are only two possible bits of information and all eight employees need one unique encoding, we cannot send a message specifying which employee got the trip since there aren't enough options in the encodings to go round. An encoding of at least three bits is needed to give each employee a unique encoding.

Dretske's *semantic theory of information* defines the notion of content to be compatible with Shannon's information theory, and his notions have gained some traction within the philosophical community [8].[1] To him, the content of a message and the amount of information in message – the number of bits an encoding would require – are different, for "saying 'There is a gnu in my backyard' does not have more content than the utterance 'There is a dog in my backyard' since the former is, statistically, less probable" [8]. According to Shannon, there is more information in the former case precisely because it is less likely than the latter [8]. So while information that is less frequent may require a larger number of bits in encoding, the content of information should be viewed as to some extent separable if compatible with Shannon's information theory, since otherwise one is led to the "absurd view that among competent speakers of language, gibberish has more meaning than semantic discourse because it is much more less frequent" [8]. Is there a way to precisely define the content of a message? Dretske defines the content of information as "a signal r carries the information that s is F when the conditional probability of s's being F, given r (and k) is 1 (but, given k alone, less than 1). k is the knowledge of the receiver" [8]. To simplify, the ***content*** of any information-bearing message is whatever is held in common between the source and the receiver as a result of the conveyance of a particular message. While this is similar to our definition of information itself, it is different. Information can measure the total in common between a source and receiver *simpliciter*. For example, two non-local humans can share quite a lot in common, and so share information, despite never having conveyed a message between each other. The content is whatever is shared in common as a causal *result* of a particular message, such as the conveyance of sentence 'Ralph won a ticket to Paris to visit the Eiffel Tower.'

In our example, the message that 'Ralph won a plane ticket to Paris to visit the Eiffel Tower' can be encoded in two different languages and still have the same relationship to content. The relationship of an encoding to its content is an ***interpretation***. The interpretation - usually via some interpreting agent be it either man or machine - 'fills' in the necessary background left out of the encoding, and maps the encoding to some content. In our previous example using binary digits as an encoding scheme, a mapping could be made between the encoding `001` to the content

[1] For an empirical justification of basing our work on Dretske's work, note that Dretske has more than a magnitude more citations than Floridi.

of Ralph while the encoding `010` could be mapped to the content of Sandro. The content of a particular message depends very much on the encoding scheme used by the interpreter. For example, one can interpret the encoding `11` as either the number eleven in the decimal encoding scheme, or the number three in the binary encoding scheme. Unlike many others, including Dretske, we shall make no claims about the nature of information, interpretation, and truth, in particular if what appears to be 'false' information is really misinformation or pseudo-information. This opens the door to the possibility of a sender sending an encoded message to a receiver that lacks the necessary capacity or resources of the receiver to decode it in the traditional paradigm of communication. The encoding would not then have an interpretation to content. This would be the standard definition of ***data***, which is information without an interpretation. One example would be if the message from Daniel that Ralph had won the plane ticket had been delivered via e-mail in French. A non-French speaker could have been aware of some very limited aspects of the e-mail (such as the time sent and the sender), but she would lack the necessary knowledge of French to decode the message's content and so to have an interpretation of the message. These terms are all illustrated in Figure 1. A source is sending a receiver a message. The information-bearing message realizes some particular encoding such as a few sentences in English and a picture of the Eiffel Tower, and the content of the message can be interpreted to be about the Eiffel Tower.

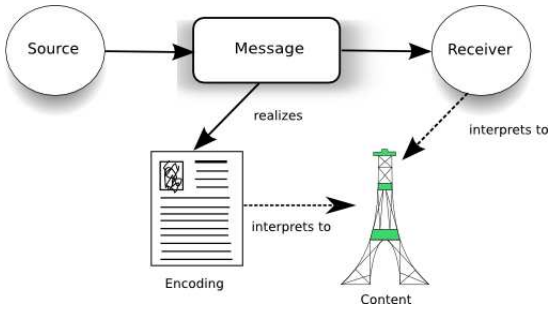

Fig. 1 Information, Encoding, Content

4 Digitality

One of the defining characteristics of information is that it can be digital, bits and bytes being shipped around by various protocols. However, we tend to know if something is digital when we spot it, and we can build digital devices, but developing an encompassing notion of digitality is a difficult task, whose solution we can only sketch here. One philosophical essay that comes surprisingly close to defining a notion of digitality is Nelson Goodman's *Languages of Art*: Given some physically

distinguishable marks, which could compose an encoding, Goodman [12] defined marks as "***finitely differentiable***" when it is possible to determine for any given mark whether it is identical to another mark or marks. This can be considered equivalent to how in categorical perception, despite variation in handwriting, a person perceives hand-written letters as being from a finite alphabet. So, equivalence classes of marks can be thought of as an application of the philosophical notion of types. This seems close to 'digital,' so that given a number of types of content in a language, a system is digital if any mark of the encoding can be interpreted to a one and only one type of content. Therefore, in between any two types of content or encoding there can not be an infinite number of other types. Digital systems are the opposite of Bateson's famous definition of information: Being digital is simply having a difference that does not make difference [1]. This is not to say there are characteristics of a mark which do not reflect its assignment in a type, and these are precisely the characteristics which are lost in digital systems. So in an analog system, every difference in some mark makes a difference, since between any two types there is another type that subsumes a unique characteristic of the token. In this manner, the prototypical digital system is the discrete distribution of integers, while the continuous numbers are the analog system par excellence, since between any real number there is another real number. The digital should include more: sentences in a language that can be realized by sound-waves or the text in an e-mail message that can be re-encoded as bits, and then this encoding realized by a series of voltages. Since the content of the information can be captured perfectly by the particulars of the encoding, this digital encoding can thus can be copied safely and effectively, just as an e-mail message can be sent many times or a digital image can be reproduced countlessly.

Lewis took aim at Goodman's interpretation of digitality in terms of determinism by arguing that digitality was actually a way to represent possibly continuous systems using the combinatorics of discrete digital states [20]. To take a less literal example, discrete mathematics can represent continuous subject matters. This insight caused Haugeland to point out that digital systems are always abstractions built on top of analog systems [16]. Haugeland further reveals the purpose of digitality to be "a mundane engineering notion, root and branch. It only makes sense as a practical means to cope with the vagarities and vicissitudes, the noise and drift, of earthy existence" [16]. Yet Haugeland does not tell us what digitality actually is, although he tells us what it does, and so it is unclear why certain systems like computers have been wildly successful due to their digitally (as in the success of analog computers was not so widespread), while others like 'integer personality ratings' have not been as successful. Without a coherent definition of digitality, it is impossible to even in principle answer questions like whether or not digitality is purely subjective [22].

Rather than fall into idealistic subjectivity, we hold that certain physical processes have the objective and material potential to be digital *if* interpreted in a particular manner - and so while interpretation does matter, it is constrained by the encoding present. Note that different interpreters can interpret the same physical encoding as 'digital' in different ways, as the marks "11" can mean eleven in decimal and three in binary notation. There are multiple ways one can state a system

is digital since digitality is a convergence between a kind of interpretation and an encoding that physically implements a correspondence between the possible states of the message and discrete types of content. So something can only be digital when content is taken into account: digitality can be defined as a relationship from an encoding to content where the encoding is finitely differentiable and the type of the encoding determines the content. In order to distinguish these types in the encoding that uphold digitality, there must be some physical regularity that serves as a *boundary* that is upheld by the physical structure of the message. When reading letters in a book, the forms of the letters serve as the boundary, not any minor variations in the quality of the printing – these analog details are left out of our interpretation. If we attempt to use an analog encoding, such as writing letters in water, the physical substrate does not have the proper physical characteristics so that digitality seems to elude us.

To implement a digital system, there must be a small chance that the system can be considered to be in a boundary state that is not part of the discrete types given by the encoding. The regularities that compose the physical boundary allows within a margin of error a discrete boundary decision to be made in the interpretation of the encoding. So, a system is capable of upholding digitality if that buffer created by the margin of error has an infinitesimal chance at any given time of being in a state that is not part of the encoding's discrete state. For example, the hands on a clock can be on the precise boundary between the markings on the clock, just not for very long. In a digital system, on a given level of abstraction, the margin of error does not propagate upwards to other levels of abstraction that supervene on the earlier level of abstractions. This first level of abstraction is 'first-order' digital, and other latter levels can be 'higher-order' digital. First-order digital created from analog physics, as we have outlined earlier, and of course higher-order digital systems can be created on top of lower-order digital systems. Although in a discrete interpretation, the encoding must be finitely differentiable, the content – as interpreted by an agent – does not have to be capable of being divided into a finite number of discrete types. For example, the encoding `00` could map to the content "Any human except Ralph or Sandro." Or, in order to capture apparently analog music stored in a digital format, one should sample the wavelength twice as often as the highest frequency of the waveform, and this leads the human to have an analog experience of the music when the music is interpreted by their stereo. So, higher-order analog can be built on top of lower-order digital systems. Furthermore, digital systems are based on our pre-digital world. This is no small achievement: We can create physical substrata that have low probabilities of being in states that do not discretely map to content at a given level of abstraction. As put by Turing, "The digital computers ... may be classified amongst the "discrete state machines," these are the machines which move by sudden jumps or clicks from one quite definite state to another. These states are sufficiently different for the possibility of confusion between them to be ignored. Strictly speaking there are no such machines. Everything really moves continuously" [28]. While "the world as we sense it on the human scale is basically analog" [18], the vast proliferation of digital technologies is possible because there are physical substrata, some more so than others, which give us the advantages that

Haugeland rightfully points out is the purpose of the digital: flawless copying and perfect reliability in a flawed and imperfect world [16].

5 Representations

Content matters! Content can be local, as when a message between two computers to 'display these bytes on the screen can translate these bytes to the screen directly without any worry about what those bytes represent to a human user. However, the content of the message may involve some distal components, such as the string "Ralph won a ticket to the Eiffel Tower in Paris," which refers things like the Eiffel Tower outside of causal reach of the computer. Any encoding of information that has non-local content is called a ***representation***. Representations are then a subset of information, and inherit the characteristics outlined of all information, such as having one or more possible encodings. This strikes to the heart of intentionality: to have some relationship to a thing that one is disconnected from is to be *about* something else. Generally, the relationship of a thing to another thing to which one is immediately causally disconnected is a ***intentional*** relationship of ***reference*** to a ***referent*** or ***referents***, the distal thing or things referred to by a representation. The thing which refers to the referent(s) we call the 'representation,' and take this to be equivalent to being a *symbol*. Yet there is a great looming contradiction: if the content is whatever is held in common between the source and the receiver as a result of the conveyance of a particular message, then how can the source and receiver share some information they are disconnected from?

We will have to make a somewhat convoluted trek to resolve this paradox. The very idea of representation is usually left under-defined as a "standing-in" intuition, so that a representation is such by virtue of "standing-in" for its referent [17]. The classic definition of a symbol from the Physical Symbol Systems Hypothesis is the genesis of this intuition regarding representations [23]: "An entity X designates an entity Y relative to a process P, if, when P takes X as input, its behavior depends on Y." There are two subtleties to Newell's definition. Firstly, the notion of a representation is grounded in the behaviour of an agent. So, what precisely counts as a representation is never context-free, but dependent upon the agent completing some action in lieu of interpreting the representation. Second, the representation *simulates* its referent, and so the representation must be local to an agent while the referent may be non-local: "This is the symbolic aspect, that having X (the symbol) is tantamount to having Y (the thing designated) for the purposes of process P" [23]. We will call X a representation, Y the *referent* of the representation, a process P the representation-using *agent*. This definition does not seem to help us in our goal of avoiding physical spookiness, since it pre-supposes a strangely Cartesian dichotomy between the referent and its representation. To the extent that this distinction is held a priori, then it is physically spooky, as it seems to require the referent and representation to somehow magically line up in order for the representation to serve as a substitute for its missing referent.

The only way to escape this trap is to give a non-spooky theory of how representations arise from referents. Brian Cantwell Smith tackles this challenge by developing a theory of representations that explains how they arise temporally [27]. Imagine Ralph finally gets to Paris and is trying to get to the Eiffel Tower. In the distance, Ralph sees the Eiffel Tower. At that very moment, Ralph and the Eiffel Tower are both physically connected via light-rays. At the moment of tracking, connected as they are by light, Ralph, its light cone, and the Eiffel Tower are a system, not distinct individuals. An alien visitor might even think they were a single individual, a 'Ralph-Eiffel Tower' system. While walking towards the Eiffel Tower, when the Eiffel Tower disappears from view (such as from being too close to it and having the view blocked by other buildings), Ralph keeps staring into the horizon, focused not on the point the Eiffel Tower was at before it went out of view, but the point where he thinks the Eiffel Tower would be, given his own walking towards it. Only when parts of the physical world, Ralph and the Eiffel Tower, are now physically separated can the agent then use a representation, such as the case of Ralph using an internal "mental image" of the Eiffel Tower to direct his walking towards it, even though he cannot see it. The agent is distinguished from the referent of its representation by virtue of not only disconnection but by the agent's attempt to track the referent, "a long-distance coupling against all the laws of physics" [27]. The local physical processes used to track the object by the subject are the representation. This notion of representation is independent of the representation being either internal or external to the particular agent, regardless of how one defines these boundaries.[2] Imagine that Ralph had been to the Eiffel Tower once before. He could have marked its location on a piece of paper by scribbling a small map. Then, the marking on the map could help guide him back as the Eiffel Tower disappears behind other buildings in the distance. Any definition of representation worth its salt should be capable of including 'external' representations, which are just as, if not more important than, the possibility of the existence of internal representations implemented neurally. Instead of positing a connection between a referent and a representation a priori, representations are introduced as products of a temporal process. This process is non-spooky since the entire process is capable of being grounded out in physical causation without any spooky action at a distance. To be grounded out in physics, all changes must be given in terms of connection in space and time. Representations are "a way of exploiting local freedom or slop in order to establish coordination with what is beyond effective reach" [27]. In order to clarify Smith's story and improve the definition of the Physical Symbol Systems Hypothesis, we consider Smith's theory of the "origin of objects" to be a ***representational cycle*** with distinct stages [14]:

- **Presentation**: Process *S* is connected with process *O*.
- **Input**: The process *S* is connected with *R*. Some local connection of *S* puts *R* in some causal relationship with process *O* via an encoding. This is entirely non-spooky since *S* and *O* are both connected with *R*. *R* eventually becomes the representation.

[2] The defining of "external" and "internal" boundaries is actually non-trivial, as shown in earlier work[15].

- **Separation**: Processes *O* and *S* change in such a way that the processes are disconnected.
- **Output**: Due to some local change in process *S*, *S* uses its connection with *R* to initiate local meaningful behavior that is in part caused by *R*.[3]

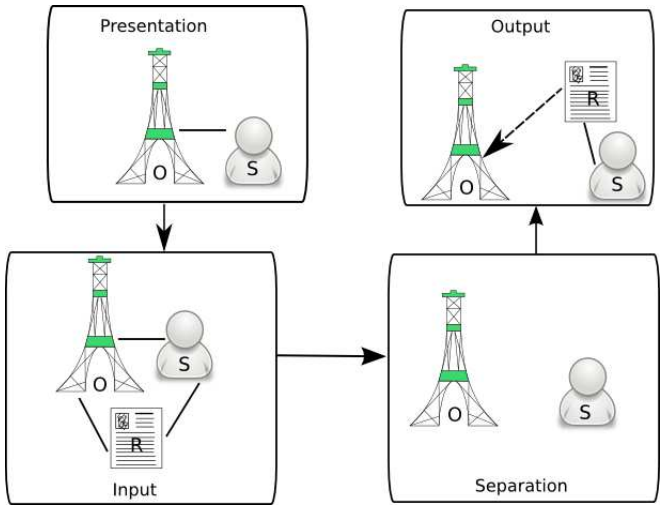

Fig. 2 The Representational Cycle

In the 'input' stage, the *referent* is the cause of some characteristic(s) of the information. The relationship of *reference* is the relationship between the encoding of the information (the representation) and the referent. The relationship of interpretation becomes one of reference when the distal aspects of the content are crucial for the meaningful behavior of the agent, as given by the 'output' stage. This is pure behaviorism insofar as the behavior may simply be impact on the cognitive structure of the agent, not necessarily 'observable' behavioral responses. So we have constructed an ability to talk about representations and reference while not presupposing that behavior depends on internal representations or that representations exist a priori at all. Representations are only needed when the relevant intelligent behavior requires some sort of co-ordination with a non-local thing. In this manner, the intentional status of representations can then be defined as the interpretation of a representation to a referent(s). This would make our notion of representation susceptible to being labeled a *correspondence theory of truth* [26], where a representation refers by some sort of structural correspondence to some referent. However, our notion of representation is much weaker, requiring only a causal history between the referent and the representation - and not just any causal relationship (since those would be nearly infinite!), but one that changes the behavior of interpreting agent as a result

[3] In terms of Newell's earlier definition, 0 is *X* while *S* is *P* and *R* is *Y*.

of the interpretation of the representation. This is opposed to some tighter notion of correspondence such as some structural 'isomorphism' between a representation and its referent [6].

The interpretation of representations should therefore not be viewed as mapping to referents, but a mapping to some content where that content leads to meaningful behavior precisely because the content is non-local. Up until now, it has been implicitly assumed that the referent is some physical entity that is non-local to the representation, but the physical entity was still existent, such as the Eiffel Tower. However, remember that the definition of non-local includes *anything* the representation is disconnected from, and so includes physical entities that may exist in the past or the future. The existence of a representation does not imply the existence of the referent or the direct acquaintance of the referent by the agent using a representation – a representation only implies that some crucial aspect of the content is non-local. However, this seems to contradict our 'input' stage in the representational cycle, which implies that part of our definition of representation is historical: for every *re*-presentation there must be a presentation, an encounter with the thing presented. By these conditions, the famous example of Putnam's ant tracing a picture of Winston Churchill by sheer accident in the sand would not count as a representation [24]. If Ralph didn't know where the Eiffel Tower was, but navigated the streets of Paris and found the Eiffel Tower by reference to a tracing of a Kandinsky painting in his notebook, then Ralph would not then be engaged in any representation-dependent meaningful behavior, since the Kandinsky painting lacks the initial presentation with the Eiffel Tower. The presentation does not have to be done by the subject that encountered the thing directly. However, the definition of a representation does not mean that the *same* agent using the representation had to be the agent with the original presentation. A representation that is created by one agent in the presence of a referent can be used by another agent as a 'stand-in' for that referent if the second agent shares the same interpretation from encoding to distal content. So, instead of relying on his own vision, Ralph buys a map and so relies on the 'second-order' representation of the map-maker, who has some historical connection to someone who actually traveled the streets of Paris and figured out where the Eiffel Tower was. One can obviously refer to Gustave Eiffel even though he is long dead and buried, and so no longer exists. Also, the referent of a representation may be a concept, like the concept of a horse, unicorns and other imaginary things, referents to future states such as 'see you next year,' and descriptive phrases whose supposed *exact* referent is unknown, such as 'the longest hair on your head on your next birthday.'

One could claim that the Eiffel Tower is simply the wrong kind of content one should be worried about as regards representation, and that one should rather be concerned with more exotic examples of infinitaryy objects such as $\aleph_1$. We would counter that it is precisely the ordinariness of the Eiffel Tower that is more important, as we can follow Clark's line that the more exotic kinds of representations descend from capabilities of abstraction developed out of sensory-motor apparatus and memory evolved in dealing with ordinary objects like the Eiffel Tower [4] - and any scientifically minded philosopher would have a hard time arguing the reverse,

namely that the ability to represent infinitary objects like $\aleph_1$ somehow evolutionarily preceded the ability to represent more mundane objects like the Eiffel Tower. The Eiffel Tower example also is actually necessary for, rather than superseded by, any supposed 'simulation' theory of representation [13]. After all, the very concept of simulation only works if there is a world to simulate. In the case, the spatio-temporally distal object the Eiffel Tower is exactly necessary to have some kind of causal (perhaps via an historical chain, one even spread out over evolutionary time) relationship to the simulation itself, the presentation implicit in any representation.

6 Conclusion

As digitality can be thought of as a convergence between the encoding and content of information, and representations as information with a non-local content, the once-insurmountable problem of digital representations then becomes rather simple: digital representations are merely digital information with non-local content. Taking as a starting point the purely causal representational cycle, a purely materialist reading of digital representations is then possible. If we identify embodiment with a certain reductive materialism, then this story lets digital representations be reconciled with embodiment. Thus, we hope our goal has the fear from certain advocates of embodiment that somehow digital representations are at their core non-materialist and anti-scientific, much less metaphysically implausible. Yet, we should also be aware of the limitations of this story we have sketched here about digitality and representations; namely this is simply a sketch to serve as what Dennett would call an "intuition pump" for a much larger story that we can hardly do justice to at this stage [7]. Massive amounts of empirical evidence needs to be gathered before we can understand the myriad possible couplings between digitality and our intuitions regarding a primarily pre-digital world, as well as the delicate intertwining of representations and our presence in the world, and a million other questions besides. Without a doubt, a much more thorough analytic argument can and should be both proposed and empirically tested. Yet without such a guiding definitional sketch as presented here, such an analysis are, such an endeavor would be mired in a confusing Tower of Babel of differing terminology and intuitions that seek to eliminate each other on metaphysical grounds.

There is a latent contradiction which we did not solve that requires further work: namely, as representations are defined by *separation* over time and space, the inexorable trajectory of computation in the era of the Internet is to eliminate this very division of time and space. The cycles of representation become ever more infinitesimal as the Internet interconnects referents ever closer with their representations. At a certain point, the operative question becomes whether or not the representation simply becomes a new kind of first-class object?[4] In other words, the ontology of the world is dynamic, created as an enactment between a multiplicity of referents

[4] This is distinctly opposed to the viewpoint of certain post-structuralist or postmodern theorists like Baudrillard that hold that representations are 'copies' that are just as real or true as their

and representations that alter each other in turn. A representation of an object is the *spreading out* of an object in time and space. It is not to say that the representational cycle and its vocabulary of referents disappear, but that they are mediated by objective sense and that the formation of a representation is just the first step of the unfolding of a new kind of object. In such a dialectic, the map becomes the territory. With the advent of digital technologies, not only the map becomes digital, but the territory itself. This points out a certain radical notion that dooms all semantic theories of information, namely that representations are not mere mirrors of the world, but representations are ontologically disruptive in of themselves. Merely semantic theories of information punt on the difficult questions of metaphysics and ontology, yet what we find in our increasingly digital and representational world is that such questions are now pressing upon us with such force that we ignore them at our own peril.

Acknowledgements This paper is an edited and improved version of a chapter of my Ph.D. dissertation entitled 'Sense and Reference on the Web," which would have been impossible without the guidance of my advisors Andy Clark and Henry S. Thompson. Also, I would like to thank Brian Cantwell Smith not only for the overall theoretical picture in his class-notes (which I have surely developed in a less able manner than he would be capable of, although also in a more strictly materialist manner than he would be comfortable with), but also for several clarifications as regards this paper. The feedback of an anonymous reviewer, as well as Vincent Mueller's encouragement and patience, have been vital.

References

1. Bateson, G.: Steps to an Ecology of Mind. University of Chicago Press, Chicago, Illinois, USA (2001)
2. Brooks, R.: Intelligence without representation. Artificial Intelligence **47**(1-3), 139–159 (1991)
3. Clark, A.: Being There: Putting Brain, Body, and World Together Again. MIT Press, Cambridge, MA (1997)
4. Clark, A.: Minds, brains, tools. In: H. Clapin (ed.) Philosophy of Mental Representation, pp. 66–90. Clarendon Press, Oxford, United Kingdom (2002)
5. Clark, A., Chalmers, D.: The extended mind. Analysis **58**(1), 7–19 (1998)
6. Cummins, R.: Representations, Targets, and Attitudes. MIT Press, Cambridge, Massachusetts, USA (1996)
7. Dennett, D.: Brainstorms: Philosophical Essays on Mind and Psychology. MIT Press, Cambridge, Massachusetts, USA (1981)
8. Dretske, F.: Knowledge and the Flow of Information. MIT Press, Cambridge, Massachusetts, USA (1981)
9. Dreyfus, H.: What Computers Still Can't Do: A critique of artificial reason. MIT Press, Cambridge, Massachusetts, USA (1979)
10. Floridi, L.: Open problems in the philosophy of information. Metaphilosophy **35**(4), 554–582 (2004)
11. Fredkin, E.: An introduction to digital philosophy. International Journal of Theoretical Physics **42**(1), 189–247 (2003)

original referent. Instead, we challenge this belief in a singularly real or authentic (and so static) ontology by incorporating the referent and representation into a new ontological object.

12. Goodman, N.: Languages of Art: An Approach to a Theory of Symbols. Bobbs-Merrill, Indianapolis, Indiana, USA (1968)
13. Grush, R.: In defence of some Cartesian assumptions concerning the brain and its operation. Biology and Philosophy (18), 53–93 (2003)
14. Halpin, H.: Representationalism: The hard problem for artificial life. In: Proceedings of Artificial Life X, pp. 527–534. Bloomington, Indiana (2006)
15. Halpin, H.: Foundations of a philosophy of collective intelligence. In: Proceedings of Convention for the Society for the Study of Artificial Intelligence and Simulation of Behavior (2008)
16. Haugeland, J.: Analog and analog. In: Mind, Brain, and Function, pp. 213–226. Harvester Press, New York City, New York, USA (1981)
17. Haugeland, J.: Representational genera. In: Philosphy and Connectionist Theory, pp. 61–89. Erlbaum, Mahwah, New Jersey, USA (1991)
18. Hayles, N.K.: My Mother was a Computer: Digital Subjects and Literary Texts. University of Chicago Press, Chicago, Illinois (2005)
19. Israel, D., Perry, J.: What is information? In: P. Hanson (ed.) Information, Language, and Cognition, pp. 1–19. University of British Columbia Press, Vancouver, Canada (1990)
20. Lewis, D.: Analog and digital. Nous **1**(5), 321–327 (1971)
21. Maturana, H., Varela, F.: Autopoiesis and Cognition: The Realization of the Living. D. Reidel Publishing, Dordecht (1973)
22. Mueller, V.: Representation in digital systems. In: Proceedings of Adapation and Representation (2007). Http://www.interdisciplines.org/adaptation/papers/7 (Last accessed March 8th 2008)
23. Newell, A.: Physical symbol systems. Cognitive Science **1**(4), 135–183 (1980)
24. Putnam, H.: The meaning of meaning. In: K. Gunderson (ed.) Language, Mind, and Knowledge. University of Minnesota Press, Minneapolis, Minnesota, USA (1975)
25. Shannon, C., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press (1963). Republished 1963
26. Smith, B.C.: The correspondence continuum. In: Proceedings of the Sixth Canadian Conference on Artificial Intelligence. Montreal, Canada (1986)
27. Smith, B.C.: The Origin of Objects. MIT Press, Cambridge, MA (1995)
28. Turing, A.M.: Computing machinery and intelligence. Mind **59**, 433–460 (1950)
29. Wheeler, M.: Reconstructing the Cognitive World: The Next Step. MIT Press, Cambridge, Massachusetts, USA (2005)