Eigenforms, Interfaces and Holographic Encoding

Toward an Evolutionary Account of Objects and Spacetime

Chris Fields • Independent Scholar • fieldsres/at/gmail.com
Donald D. Hoffman • University of California, Irvine, USA • ddhoff/at/uci.edu
Chetan Prakash • California State University, USA • cprakash/at/csusb.edu
Robert Prentner • ETH Zurich, Switzerland • robert.prentner/at/phil.gess.ethz.ch

> Context • The evolution of perceptual systems and hence of observers remains largely disconnected from the question of the emergence of classical objects and spacetime. This disconnection between the biosciences and physics impedes progress toward understanding the role of the "observer" in physical theory. **> Problem** • In this article we consider the problem of how to understand objects and spacetime in observer-relative evolutionary terms. **> Method** • We rely on a comparative analysis using multiple formal frameworks. **> Results** • The eigenform construct of von Foerster is compared to other formal representations of observer–environment interactions. Eigenforms are shown to be encoded on observer-environment interfaces and to encode fitness consequences of actions. Space and time are components of observational outcomes in this framework; it is suggested that spacetime constitutes an error-correcting code for fitness consequences. **> Implications** • Our results contribute to an understanding of the world in which neither objects nor spacetime are observer-independent. **> Constructivist content** • The eigenform concept of von Foerster is linked to the concepts of decoherence and holographic encoding from physics and the concept of fitness from evolutionary biology.

> Key words • **Active inference, boundary, conscious agent, icon, Markov blanket, redundancy.**

Introduction

« 1 » Heinz von Foerster (1976) introduced the eigenform and eigenbehavior concepts by considering an agent that both observes and acts on a surrounding world: an *eigenform* is an observation that remains invariant, in the limit of long interaction time, under some class of behaviors, while an *eigenbehavior* is an action that, in the same limit, leaves some eigenform invariant. These concepts naturally suggest an abstract picture in which the eigenbehavior continually reproduces the eigenform, independently of any other features or dynamics of the world. In this picture, eigenform and eigenbehavior compose a single reflexive system; all other aspects of the world can be neglected. Louis Kauffman has shown, conversely, that all such reflexive systems have eigenforms and eigenbehaviors as invariants. Kauffman elevates the reflexivity of such self-reproducing eigenform-eigenbehavior systems to a principle of cosmology:

> "The Universe is constructed in such a way that it can refer to itself […] the universe can pretend that it is two and then let itself refer to the two, and find that it has in the process referred only to the one, that is, itself." (Kauffman 2009: 134)

This formulation makes explicit an important point: that there is no difference in *substance*, and hence no metaphysical dualism, between agent and environment.

« 2 » Here we pursue the notion of an eigenform not from the perspective of an abstract reflexive system, but rather from von Foerster's original perspective of an agent that observes and acts on its world, a world that can be taken to be the rest of the Universe in which the agent is embedded. We impose, in other words, an "epistemic cut" in the sense used by John von Neumann (1955) or Howard Pattee (2001) between agent and world for the purposes of theory construction. It is from this perspective that an eigenform becomes, or perhaps better, *serves as* an object that the agent observes and acts with respect to. This agent-centered perspective, when combined with the essential external perspective of the theorist, allows us to consider the ecological situation of an agent for whom every observation presents multiple objects, every object allows multiple actions, and every pairing of an object with an action has consequences that may be good or bad for the agent. We compare the description of this situation in terms of eigenforms to its description in two independently developed formal representations of the agent-world interaction: the conscious agent formalism of Donald Hoffman and Chetan Prakash (2014) and the Markov blanket formalism of Judea Pearl (1988) as applied to biological systems by Karl Friston

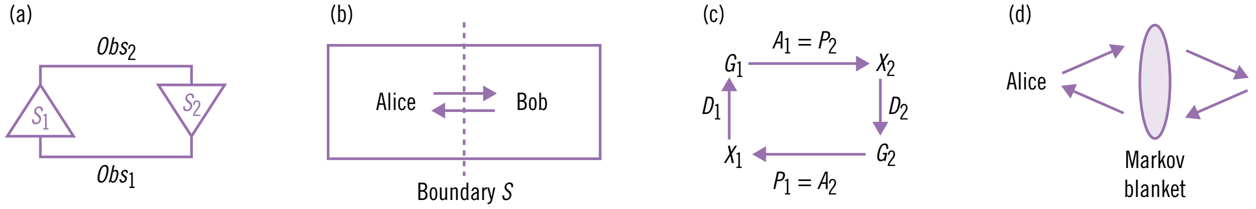


Figure 1 • Four representations of two-agent, or alternatively, agent-environment interaction. (a) Two agents S_1 and S_2, here depicted as computational processes, exchange observations Obs_1 and Obs_2 (adapted from Foerster 1976: 94). (b) Two agents, or alternatively two classical black boxes, Alice and Bob exchange inputs and outputs across a boundary S that is in principle arbitrarily movable as described in Fields (2016). Alice's outputs are Bob's inputs and vice versa. (c) Two conscious agents as defined by Hoffman and Prakash (2014) act on each other. Here X_1 and G_1 and X_2 and G_2 are measurable spaces representing the experiences and available actions, respectively, of the two agents; D_1 and D_2, P_1 and P_2, and A_2 and A_2 are Markov kernels representing the decision processes, perceptions, and executed actions, respectively, of the two agents. (d) Two agents interact via an intervening Markov blanket as described in Friston (2013). Arrows represent Markov processes.

(2013). In both of these latter representations, the agent's observations and actions "pass through" a boundary or *interface* that separates the agent – even if this separation is purely notional – from its observed world. We show that eigenforms can be regarded as "icons" specifying possible interactions that are encoded on this interface. We then suggest that this notion of an encoding of information about possible interactions on an interface is in fact very general, by showing that it corresponds to the notion of holography developed within quantum information theory. In this case, the encoding can be regarded as "recorded" by the process of quantum decoherence, confirming the close relationship between the eigenform concept and quantum theory already suggested by Kauffman (2003, 2011).

« 3 » Considering eigenforms as encodings of information *for* a particular agent on that agent's interface with its observed world allows us to ask what information an eigenform encodes. If perceived "objects" are tokens for eigenforms, what is their informational role? The Interface Theory of Perception (ITP) of Hoffman, Manish Singh and Prakash (2015) provides a *prima facie* surprising answer: that "objects" do not encode information about the ontological or causal structure of the world, but rather information about the structure of the fitness function that relates the agent to the world. This information is object-relative, but not object-specific: an interaction with one object can have fitness consequences that affect interactions with other objects. An eigenform, in other words, encodes information not just about its own stability, but also about the stability of other eigenforms. What kind of encoding, we then ask, can have this property? We suggest that spacetime itself, including both the space in which objects appear to be embedded and the time over which they appear to persist, is a relational, error-correcting code for the fitness consequences of interactions. The forms and locations of "objects" in "space" encode probabilistic information about what future interactions with these or other objects, if they occur at all, may be like. The persistence of an "object" in "time" encodes the robustness of the corresponding eigenform as an attractor. Eigenforms have evolved, we argue, to make this encoding of future consequences as precise as possible given the energetic and other resource constraints of the encoding interface.

The interface

« 4 » As von Foerster recognized, a reflexive model escapes solipsism when the "world" or "environment" of each agent includes other agents, or in the limit *is* another agent (e.g., Foerster 1960). Such a two-agent model is shown in Figure 1a; here two agents S_1 and S_2 exchange observations Obs_1 and Obs_2 (Foerster 1976: 94). From the perspective of either agent, the other agent is its entire "world" and every observation appears to be an observation of this entire world; there is nothing else with which the agent interacts, and hence nothing else that it observes. It is only from the perspective of a theorist describing the overall situation "from the outside" that the two agents and their exchange of observations within the closed-loop system can be made explicit.

« 5 » The closed-loop, two-agent exchange in Figure 1a involves an apparent paradox: each agent receives information from the other, so the total information in a two-agent system appears to increase. Any such increase in a closed system, as von Foerster (1960) notes, appears to violate the 2nd law of thermodynamics. Indeed, any agent, as a self-organizing system, must "eat energy and order from its environment" (Foerster 1960: 36) in order to survive; from the perspective of any such agent, the order in its environment must decrease as it is "eaten." The environment of either agent in Figure 1a is the other agent; hence each agent must perceive the other as *losing* information. It is here that the difference between the agents' and the theorist's perspectives becomes critically important. As Max Tegmark (2012) remarks in a similar context, neither agent has observational access to the total entropy of the two-agent system (neither agent has the theorist's perspective); neither agent can get "outside" the system to measure the total entropy. The total entropy of the two-agent system could be zero, as indeed it would be if the agents were quantum-mechanical systems with an entangled joint state (in this case, each agent would see itself communicating, but an outside observer would see no communication as discussed further below). It is only the agents' principled lack of observational access to the system in which they are embedded that allows each agent to consider itself to be gaining information at the expense of its environment. Hence the second law is respected from each agent's individual perspective. This comports well with the probabilities that appear in the second law's being subjective, not objective.

« 6 » The lack of observational access that rescues Figure 1a from paradox has a second important consequence: the environment of each agent becomes a classical black box, a system to which observers have only external access. More formally, a classical black box is a system about which no observer can have more (non-hypothetical) information than is contained in a finite list of finite-length bit strings representing observed input-output transitions (Ashby 1956; for a recent review, see Fields 2016a). Because neither agent can see "inside" the black box of its environment – this is, after all, what "no observational access" *means* – neither agent knows what its environment contains. The two agents of Figure 1a can, therefore, also be represented as two interacting black boxes; we give them their traditional names Alice and Bob (Figure 1b; cf. the similar construction of Ranulph Glanville 1982: Figure 5, where the theorist's perspective is made explicit). Alice gives inputs to the unknown system Bob and receives outputs in return; the situation is the same from Bob's point of view. Edward Moore's (1956) theorem assures that neither Alice nor Bob can determine the complete state space or dynamics of the other from finite input-output observations (see Fields 2013, 2016 for extensive discussion). Either must, therefore, regard the other as a "non-trivial machine," i.e., as a system whose behavior is unpredictable in principle as von Foerster (1973) emphasizes. Principled unpredictability is considered by some to indicate autonomy or "free will" and hence agency from the perspective of external observers (e.g., Conway & Kochen 2006; Fuchs 2010; Fields 2013); even infants associate agency with behavioral unpredictability (e.g., Luo & Baillargeon 2010; Csibra & Gergely 2012). Any black box can, on this view, be considered to be or at least contain an agent. The inability of any observer of a black box to determine where in the box an enclosed agent is, or how much of the box the enclosed agent occupies is what allows the limiting case in which the other agent *is* the box (Fields 2016a), and is hence what allows the two-agent representation in Figure 1a.

« 7 » The position of the boundary *S* separating Alice from Bob in Figure 1b is, like the total entropy of the joint Alice + Bob system, definable only from the "god's eye" perspective of the theorist. Moving the boundary changes the "sizes" of Alice and Bob and hence their definitions as "systems." It also changes what "counts" for each of them as an input or an output. However, moving the boundary *S* changes nothing about the relationship of mutual exchange between Alice and Bob, and indeed nothing about the behavior of the joint system they compose. This invariance under changes in the positions of boundaries drawn by theorists is built deeply into the formalisms of both classical and quantum physics (Fields 2016a); it is, indeed, this invariance that allows theorists to choose "systems of interest" arbitrarily. It is implicit in von Foerster's (1976) and Kauffman's (2009) reduction of the agent-environment dynamics to the reflexive dynamics of a single, unitary system. The Alice–Bob boundary being arbitrarily movable means that Alice and Bob do not know, and cannot determine, where in the joint system their mutual boundary is. Each can only locate the boundary from her or his own perspective; the "god's eye" perspective needed to locate it within the joint system is unavailable. Not only can they not observe the "interior" of their interaction partner/environment, they cannot observe the boundary separating themselves from their partner/environment. All that either Alice or Bob can observe is the sequence of "inputs" that cross their respective boundaries from their respective environments. These sequences of inputs are the totality of their perceptual, as opposed to internally generated or introspective, experiences.

« 8 » As agents, Alice and Bob not only perceive, but also act; eigenforms are fixed points of and hence encode regularities in the perception-action relationship. Why should such regularities exist? From the theorist's point of view, eigenforms are inevitable, as shown by von Foerster (1976) and made more explicit by Kauffman (2003, 2009). Such a proof does not, however, say *which* eigenforms are inevitable. From an agent's perspective, an eigenform is an eigen*percept*, a percept that does not change when the "right" action – the eigenbehavior – is executed. Such an eigenpercept has persistence over time *if* the right action is taken; the wrong action may lead to its disappearance. An autonomous agent must *choose* the right action to take in any particular circumstance, i.e., given any combination of current state and current percept. To the eigenform-eigenbehavior concept, therefore, we may add the notion of an eigen*decision*, the decision to execute the eigenbehavior that results in renewal of the eigenform. While autonomy in the non-trivial machine sense inferred above is somewhat abstract, a requirement for autonomous decision-making at least suggests an awareness of potential consequences and hence consciousness.

« 9 » A minimal formal model of a conscious agent (CA) that experiences perceptual input from the world *W* in which it is embedded, decides between possible actions to take on the basis of that input, and then executes the selected action on *W* has been developed by Hoffman and Prakash (2014), who show that this minimal model is computationally universal. They propose as the thesis of "conscious realism" that the world *W* can always be considered to itself be a CA; in this case, the agent-world interaction can be represented as in Figure 1c (adapted from Hoffman and Prakash 2014: Figure 2). Conscious realism incorporates, clearly, the assumption discussed above that the limit in which the other agent "fills" the entire environment exists. As in the case of a black-box agent, this assumption can be stated as a claim about observational access: no agent can demonstrate by observation that its environment or any component thereof is not also a conscious agent. Conscious realism makes each agent's action the other agent's perception in Figure 1c, just as they are in Figures 1a and 1b. In either agent's case, the space *X* of experiences contains all of the information on which its choices of actions, which are assumed to be autonomous and hence "free," may be based, including any memories, values, goals, or other introspectively accessible content. It is important to emphasize that a CA does not experience the operations *P*, *D* or *A*, but only the elements of the experience space *X*; an account of how experiences are "written on" *X* is discussed below.

« 10 » The analog in Figure 1c of the arbitrarily-movable inter-agent boundary S in Figure 1b is the purely notional point at which Alice's action A becomes Bob's perception P and vice versa. Consistent with the discussion above, this point is invisible. From Bob's perspective, Alice acts directly on his experience space X_{Bob}; similarly for Alice. We can, therefore, simply identify the two oriented surfaces of the boundary S, the surface facing Bob and the surface facing Alice, with the experience spaces X_{Bob} and X_{Alice} respectively. In this case, Alice and Bob each act outwardly, through their own experience spaces, on the experience space of the other. Note that making this identification of the two surfaces of S with the experience spaces X_{Bob} and X_{Alice} renders Alice and Bob neither "open" nor "closed" in the mereotopological sense (Smith 1996); Alice and Bob rather share a single boundary that "belongs" to neither of them (for further discussion of this point, see Fields 2014). Treating each agent's outward action on the other agent as experienced by the agent performing the action requires giving the space X a structure that allocates some part of X for the recording of at least short-term memories of executed actions. Recording each action as it is executed, even if this record is "forgotten" immediately thereafter, is the minimal requirement for experienced learning and hence for experientially understanding or expecting anything about the environment. It is, similarly, the minimal requirement for any experience of acting, i.e., of *being an agent*.

« 11 » The idea that interacting agents interact via a shared, epistemically impenetrable boundary has been formulated independently by Friston (2013), who provides an analog, using Pearl's (1988) Markov blanket formalism, of the von Foerster–Kauffman demonstration that eigenforms are inevitable. A Markov blanket is a collection of nodes, such that knowing the state of this collection renders the states of two sets of nodes interacting only via the blanket conditionally independent (Figure 1d). Pearl (1988) shows that a Markov blanket appears whenever a random dynamical system is factored into parts (see Friston 2013 or Friston et al. 2015 for more informal discussions). The blanket effectively encodes information about how the actions of one system affect the state of the other; it thus "translates" Alice's actions into Bob's perceptions and vice-versa, just as the boundary S does in Figure 1b. It plays the role that von Foerster (1979) assigns, in a very general sense, to language. Either agent's interactions with its own surface of the blanket can be described in terms of Bayesian "active inference," in which the agent can choose, given any percept, either to alter its expectations about the world, i.e., about the probabilities of future percepts, or to act in some way that changes the percept (Friston 2010; 2013). This conceptualization of the agent's potential responses to a percept has led to architectural predictions in both neuroscience (Adams, Friston & Bastos 2015) and developmental biology (Friston et al. 2015).

« 12 » The idea that perceptions, in the broad sense of informational inputs from the world, appear on a "surface" separating an agent from the world on which it acts – a surface that not only presents information and enables action, but also blocks further epistemic access to what is on the other side – immediately suggests a familiar analogy: the user interface of a computer. Like the surface S in Figure 1b, the user interface of a computer presents *all* of the information about the computer's internal state that the user can access without disrupting the computer's function. User interfaces provide highly abstracted representations of the computer's internal state, each of which allows a circumscribed set of possible actions. They systematically hide not just the behavioral complexity, but the entire physical and causal structure of the computer. User interfaces are, moreover, ambiguous about this structure by design: as with any virtual machine (Smith & Nair 2005), platform independence is a major component of a user interface's utility. Computer programs are by no means alone in having these properties; as Willard Van Orman Quine (1960; see also Quine 1970) points out, all human natural languages have them. If a model-theoretic approach to semantics (Tarski 1944) is adopted, all "languages" of any kind have them. A computer's user interface, however, *obviously* has them, which is what makes it a particularly good analogy.

« 13 » The Interface Theory of Perception (Hoffman, Singh & Prakash 2015) challenges the still-dominant assumption that human perception is at least approximately veridical (e.g., Marr 1982; Palmer 1999; Geisler & Diehl 2003; Trivers 2011; Pizlo, Li, Sawada & Steinman 2014) with the claim that human perception and action are interactions with a "user interface" formed of conscious experience that systematically hides both the ontology and the causal structure of the world. As stable action-perception associations, eigenforms "live on" this interface. The icons and windows of a computer interface are placed there by designers. There is, however, no "designer" in ITP. We discuss in the next section how information can be encoded on an interface by the process of information exchange itself.

Holographic encoding

« 14 » Objects as spatially bounded, temporally persistent, internally cohesive, causally independent entities are simply taken for granted as part of the "classical worldview" (roughly corresponding to what Edmund Husserl 2012 called the "natural attitude") on which human material culture is largely based. This classical conception of objecthood is so critical to ordinary human cognition that it is widely regarded as innate (e.g., Spelke 1994; Baillargeon 2008). Albert Einstein viewed the boundedness, persistence and causal independence of objects as critical to science, claiming that "without such an assumption of the mutually independent existence (the "being-thus") of spatially distant things, an assumption which originates in everyday thought, physical thought in the sense familiar to us would not be possible" (quoted in Fuchs & Stacey 2016: 6). Niels Bohr (1928, 1958) emphasized that items of laboratory apparatus must be regarded as classical objects if the notion of an "observational outcome" is to make sense. Eugene Wigner's (1962) "friend" paradox nicely illustrates the consequences, within the classical worldview, of not treating other observers as bounded, persistent objects: they not only lose any claim to consciousness and hence observerhood, they become entangled with the rest of the world and effectively disappear.

« 15 » The assumptions of epistemic transparency and objective persistence over time underlying the classical worldview

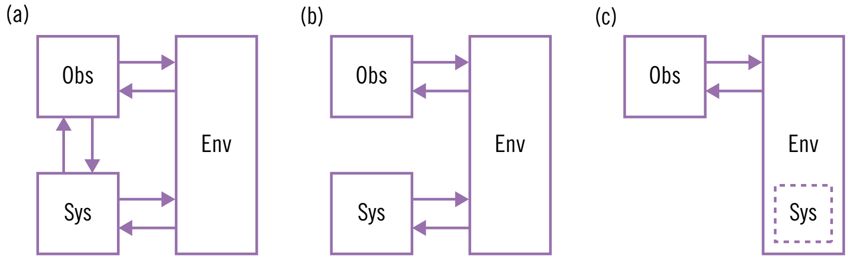


Figure 2 • Three views of decoherence. (a) An observer (Obs) prepares and measures a quantum system (Sys). Both independently interact with a large surrounding environment (Env), which renders their states effectively classical by a decoherence mechanism (e.g., ambient photon scattering). (b) The "environment as witness" formulation of Ollivier, Poulin & Zurek (2004, 2005), in which the observer interacts with the system only via the "witnessing" environment. This environment decoheres the system but interacts effectively classically with the observer. (c) If the assumption of environmental transparency is rejected, the environment becomes a black box. In this case, the system is completely embedded within it in a way that provides the observer with no access to the system-environment boundary. In this case, decoherence can only be defined at the observer-environment boundary.

have been criticized at least since Heraclitus. Quantum theory, however, forcefully raises the question of how it could even be *possible* to experience spatially bounded, temporally persistent, internally cohesive, causally independent entities. While some physicists still reject it (e.g., Ghirardi, Rimini & Weber 1986; Penrose 1996; Weinberg 2012), unitary quantum theory with no scale-dependent physical "collapse" mechanism is increasingly supported by both experiments (e.g., Eibenberger et al. 2013; Hensen et al. 2015; Manning, Khakimov, Dall & Truscott 2015; Rubino et al. 2017) and theoretical considerations (e.g., Schlosshauer 2006; Tegmark 2012; Saini & Stojkovic 2015; Susskind 2016). In unitary quantum theory, the universe is permanently in an entangled state; there *are no* classical objects. While the appearance of classicality in such a universe is given multiple explanations (for overviews, see Landsman 2007; Wallace 2008), since the 1980s most have appealed in some way to a process of *decoherence*, i.e., an apparent removal of quantum coherence that results in an apparently classical object in an apparently classical state (for reviews, see Zurek 2003; Schlosshauer 2007).

« 16 » Three views of the decoherence process are shown in Figure 2. In the original environment-induced decoherence process of Dieter Zeh (1970, 1973), an "environment" such as a macroscopic apparatus or the ambient photon field interacts continuously with both the observer and the system being observed (Figure 2a; cf. Tegmark 2012: Figure 2). This interaction effectively removes quantum coherence from both observer and system by spreading it over the many unobserved – and in practice unobservable – states of the environment (formally, the degrees of freedom of the environment are traced over). With both observer and observed system now in effectively classical states (formally, eigenstates of their respective interaction Hamiltonians with the environment), both the preparation and measurement interactions are effectively classical. As pointed out by Harold Ollivier, David Poulin and Wojciech Zurek (2004, 2005), however, observers typically interact with systems of interest only via an apparatus or an ambient field such as the photon field (Figure 2b; cf. Ollivier, Poulin & Zurek 2005: Figure 1). This intervening environment serves as a "witness" that both decoheres the system and encodes information about its state (formally, information about the eigenstates of the system-environment interaction Hamiltonian) in a way that is accessible to the observer – indeed, to multiple independent observers – via an effectively classical interaction. In this picture, the witnessing environment "does all the work" of observation; the human observers read their observational outcomes off from the environment in the same way that they would read them out of a shared or multiply copied book. While the indirectly observed "system" is quantum, the directly observed components of the environment constitute, in this case, an effectively classical object that stands between the observer and the quantum system of interest.

« 17 » The environment-as-witness formulation of decoherence assumes that the observer knows and can characterize the system-environment boundary; the intervening environment is, in other words, assumed to be at least epistemically "transparent." What happens if this assumption of a transparent environment is rejected? In this case, the environment becomes a black box. Any "systems" are contained fully within it, in such a way that their boundaries, if they have them, are observationally inaccessible (Figure 2c; cf. Fields 2016a: Figure 1). From the observer's perspective, it is completely consistent with all available observational outcomes to treat the "system" as expanding to fill the entire "environment" (formally, system and environment are in an entangled quantum state and so cannot be assigned quantum states individually); this is precisely the limiting case discussed above. If the system-environment boundary cannot be defined, however, a decoherence interaction between system and environment cannot be defined either (Fields 2012). Decoherence can, in this case, only be defined at the observer-environment boundary, i.e., at the interface characterized above. This process is illustrated in Figure 3. The quantum state Ψ "passes through" the interface to produce an observational outcome x_i. This outcome is defined at the observer-environment boundary (formally, it is an eigenvalue of the observer-environment interaction Hamiltonian). If receiving the observational outcome x_i is to have any determinate effect on the observer, e.g., if it is to be an input to a decision process that selects a next action to perform, then it must be a *classical* outcome. To characterize x_i as classical is just to say that decoherence actually happens; hence it is to say that the observer-environment interaction actually occurs from the perspectives of both observer and environment. A classical outcome can be recorded as a classical bit string, e.g., a finite sequence of binary numbers; indeed it must be recorded in a thermodynamically irreversible way if it is to be considered to have a causal effect (Landauer 1961, 1999; Bennett 2003). Where is it encoded? In the CA model, it is encoded on the space *X* of experiences. As discussed above, this space *X* can simply be identified with the interface. Hence, we can regard the classical observational outcome value xi as encoded on the interface itself, as shown in Figure 3.

« 18 » Encodings of classical information – information that can be written as a finite bit string – on surfaces at which interactions are defined are called *holographic* by physicists. Such holographic encodings were first characterized for the surfaces bounded by the event horizons of black holes (Bekenstein 1973) and were extended to the surface of the observable universe as a whole by Gerard 't Hooft (1993) and Leonard Susskind (1995). Holographic encodings record on the surface of a system all of the information that may be obtained from it by observation; to say that a system has a holographic encoding (i.e., satisfies the "holographic principle") is to say that its *observationally accessible* information content is proportional to its surface area, not to its volume (reviewed by Bousso 2002). While the terms "surface area" and "volume" here suggest ordinary three-dimensional space, the concept of holography is much more general, applying to any system with a bounding surface and an "interior" or as physicists call it, a "bulk," that is contained within the boundary. A classical black box provides a suitably abstract example. The boundary of the black box can be taken to comprise only the degrees of freedom that encode the inputs to and outputs from the box; this restricted notion of a boundary corresponds to the restricted notions of a "system" and an "observer" commonly employed in discussions of environment-induced decoherence (e.g., Tegmark 2012). In this case, the "bulk" of the black box comprises all of the non-boundary degrees of freedom, in particular, all of the degrees of freedom involved in the process of generating the next output in response to a given input. It is precisely these "bulk" degrees of freedom to which observers of a black box have no access; indeed Moore's (1956) theorem prevents them from determining any more than a lower limit on the number of bulk degrees of freedom of a black box. The amount of information that can be obtained from a black box is strictly limited by the total coding capacity of its boundary degrees of freedom. This coding capacity can be expressed precisely as an abstract dimension. Let $\{\xi_i\}$ be the set of mutually independent degrees of freedom of the boundary, and let ni be the number of possible distinct values of the *i*-th boundary degree of freedom ξ_i. The dimension of the boundary is then the sum of the numbers ni over all the degrees of freedom in $\{\xi_i\}$:

$$d_{boundary} = \sum \{\xi_i\}\, n_i$$

Similarly, let $\{\zeta_j\}$ be the set of mutually-independent degrees of freedom of the bulk, and let m_j be the number of possible distinct values of the *j*-th bulk degree of freedom ζ_j. The dimension of the bulk is then:

$$d_{bulk} = \sum \{\zeta_j\}\, m_j$$

The amount of information that an observer can obtain from any black box is clearly proportional to $d_{boundary}$, not to d_{bulk}; hence any black box satisfies the holographic principle.

« 19 » The only information that an observer can obtain about the surrounding environment is the information that can be encoded on the observer-environment interface by decoherence; the environment of any observer is, therefore, a black box and satisfies the holographic principle (cf. Fields 2016a). The loop from Figure 3 back to Figure 1b is thus closed: from the environment's perspective, the observer also satisfies the holographic principle, as the environment can only obtain information about the observer that can be encoded on the observer-environment boundary.

« 20 » Kauffman (2003, 2011) has previously related the eigenvectors representing observable degrees of freedom and eigenvalues representing observable outcome values in quantum theory to eigenforms as stable outcomes of repeated measurements. Indeed, the stability of observational outcomes under exactly repeated measurements underlies the notion of "system preparation" and is often regarded as an axiom of quantum theory (e.g., Zurek 2003: 747). The above discussion localizes this conceptual connection to the observer-environment boundary – the interface as described by ITP – and shows that the connection is implemented by decoherence, the process that creates stable classical records of transient quantum states.

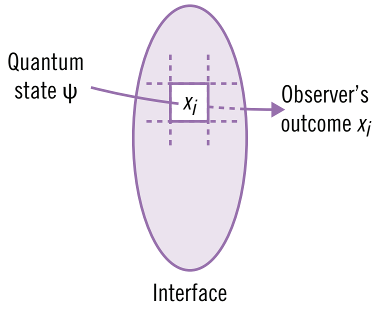


Figure 3 • Decoherence encodes a classical outcome value x_i on the observer-environment interface. Such an encoding is required if receipt of the observational outcome is to be considered to have any effect on the observer's subsequent behavior. This encoding is holographic, i.e., the only information about the environment that the observer can obtain is the information that can be encoded on the observer-environment interface by decoherence.

Interfaces encode fitness

« 21 » As discussed above, information is classical to the extent that it has an effect on decision and action, i.e., to the extent that it is *useful* to the agent that receives it. Information that has no effect – information that changes nothing about its recipient – is information that has not been recorded. As Gregory Bateson put it, "what we mean by information – the elementary unit of information – *is a difference which makes a difference*" (Bateson 1987: 460; emphasis in original). All the information that agents possess is information that has had some effect on them; it is all "pragmatic information" in Juan Roederer's (2005) sense, information that enables doing something. Von Foerster (1970) makes a similar point, quoting Jerzy Konorski: "information and its utilization are inseparable […] one single process" (Foerster 1970: 46).

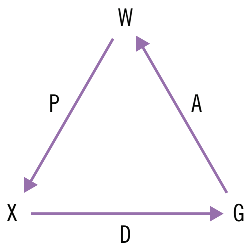


Figure 4 • A CA as defined by Hoffman & Prakash (2014) as a perceive-decide-act (*P-D-A*) loop through a "world" W, which takes the place of the "second agent" X_2-D_2-G_2 in Figure 1c.

« 22 » In the CA model of Hoffman and Prakash (2014), the recursive loop is perceive-decide-act (*P-D-A*) as shown in Figure 4. Here perceptions (*P*) come from and actions (*A*) are on the "world" *W* of the CA; *W* replaces the "second agent" X_2-D_2-G_2 in Figure 1c. A CA is defined by the continued performance of this *P-D-A* loop. Should the recursion be for any reason interrupted – should there occur a perception after which no decision follows, a decision after which no action (including the action: take no action) follows, or an action after which no perception follows – the CA ceases to exist. It is "dead."

« 23 » We can, therefore, define the *fitness* of a CA as the probability of continued recursion, and the *fitness function F* of a CA as a mapping F: $X \times G \times W \rightarrow$ Non-negative Reals. "Continued recursion" is "viability" in Ernst von Glasersfeld's (1981) sense for a CA; the CA only survives as long as its *P-D-A* loop "keeps working." The meaning of *F* becomes particularly clear when the world *W* is regarded as a second agent as in Figure 1c. The state *w* of *W* being such that, for states *x* of *X* and *g* of *G*, $F(x, g, w) = 0$ means that the world acts on the agent in a such a way that the agent cannot respond. This is a lethal action. As *W* is itself defined relative to the agent – it is *that agent's* world – *W* "dies" as well, following such an action.

« 24 » We are now in a position to see what interfaces encode. An interface encodes, by its very existence, the fact that it has not permitted a lethal action in either direction: for every triple of states (x, g, w) that has occurred so far, $F(x, g, w) > 0$. It has not, in particular, allowed an action after which no perception follows, or a perception from which no action follows. This can be expressed probabilistically: an interface encodes, by its very existence, the fact that the probabilities of lethal perceptions and actions have (at least so far) been low enough that none has occurred. The probabilities of perceptions and actions are, however, specified by the kernels *P*, *D* and *A* and the initial state (x_0, g_0, w_0). If we identify the interface with *X* as discussed above, a state *x* of *X* can be viewed as specifying a probability distribution $Prob(g'|x, g) = D(x, g; g')$ of the next state g' of *G* given the current state via the Markov kernel *D* and a probability distribution $Prob(w'|g, w) = A(g, w; w')$ of the next state w' of *W* via the kernel *A*. Here the kernel action $D(x, g; g')$ is the probability of deciding on g', given that the current percept is *x* and the previous decision was *g*; similarly for $A(g, w; w')$. From these an expected fitness $EF(x|g, w)$ can be calculated by summing over the fitness values of the future states (x, g', w') that can immediately follow the current state (x, g, w), with each future state weighted by its probability:

$$EF(x \mid g, w) = \sum\nolimits_{g'w'} F(x, g', w') \; Prob(g' \mid x, g) \, Prob(w' \mid g, w)$$

or making the operator actions explicit:

$$EF(x \mid g, w) = \sum\nolimits_{g'w'} F(x, g', w') \; D(x, g; g') \, A(g, w; w')$$

Interfaces, therefore, encode *expected fitness*. They encode their own best estimates of their likelihood of survival, i.e., their likelihood of receiving a next input and transmitting a next action.

« 25 » If interfaces encode information about fitness, then they do not encode information about the observer-independent ontology or causal structure of the world. In the present conceptual framework, of course, this is tautologous: there is no observer-independent ontology or causal structure in any world that is defined only relative to an observer. From the perspective of the classical worldview, however, this is a surprising result. It is supported by evolutionary game-theory experiments that adopt the classical worldview in so far as they assign "true" world states in an agent-independent manner, but show that agents that make decisions based on these "true" world states are generally driven to extinction by agents that make decisions solely on the basis of expected fitness (Mark, Marion & Hoffman 2010). These empirical results have since been put on a rigorous footing by a "fitness beats truth" theorem demonstrating that decision strategies based on expected fitness will dominate decision strategies based on the "truth" about the world for all but a generically small subset of fitness functions. The "fitness beats truth" theorem provides a formal justification for von Glasersfeld's remark that "we must never say that our knowledge is 'true' in the sense that it reflects an ontologically real world" (Glasersfeld 1981: 93).

« 26 » Making use of the computer interface analogy, Hoffman, Singh and Prakash (2015) characterize perceived "objects" as "icons" on an agent's interface. These icons encode "packages" of expected fitness consequences, what James Gibson (1979) called "affordances," though Gibson tended to view affordances as "objectively" encoded by the environment. An icon that is a perceived coffee cup, for example, encodes the expected fitness of its own use for drinking coffee. They are useful to the extent that they support behaviors – at least approximate eigenbehaviors – that leave their structure at least approximately constant. As noted earlier with respect to experiences of actions, stable icons representing "objects" with "identity over time" or "processes" that "unfold in time" require some components of the experience set *X* to be allocated to distinct collections of "memory" and "expectation" experiences. As limits of an infinite recursive process, as well as fixed points for that very process, eigenforms are encodings of their own fitness ($F \rightarrow \infty$ in the $t \rightarrow \infty$ limit) that the icons manipulated by finite organisms only approximate.

« 27 » It is important to note that the information about expected fitness that icons encode is non-local. Actions taken with respect to one icon can have consequences for future interactions with others; one's actions with respect to a perceived kitchen knife, for example, can have consequences for how one interacts later with a perceived computer. An agent that stops interacting, moreover, stops interacting with everything. Such non-local effects suggest apparent causal relations between the icons themselves. Causation in turn suggests an apparent spacetime in which causal processes operate. *Experienced* spacetime, however, must be encoded, like the icons themselves, on the interface. How is this done?

Spacetime as an error-correcting code

« 28 » As noted earlier, the agent-environment interface can be characterized in abstraction from any notion of ordinary three-dimensional space. Human perception, however, is resolutely spatial: the "objects" we see occupy space and move in space, and the actions we take are taken in space. Human experience, moreover, unfolds in time. Where does this spacetime come from? The recursion that gives rise to eigenforms provides a natural "counter" for time; this conception of time as an agent-specific counter for experience is built into the CA framework (Hoffman & Prakash 2014). What, however, about space? What is it about perception-action interfaces that makes them spatial, and what explains three-dimensionality?

« 29 » We suggest that space, and by extension spacetime, provides an error-correcting code for fitness consequences. A spatiotemporal encoding provides a way of "spreading out" information about fitness in a way that allows redundancy and hence an ability to detect and correct perceptual errors. To see the value of a spatial encoding, consider the information about quantity encoded by the positive whole numbers. These numbers are just discrete points on the real line, hence they can be represented simply as a sequence of points:

• • • • • • • • • • • • • • • • • •

This representation can even be compressed further:

•

« 30 » Such representations are, however, useless: there is no way to tell, for example, that "•" represents 4 while "•" represents 27. Making this distinction requires adding a spatial dimension that allows a planar character like "4" to be drawn out. This added dimension allows redundancy, as shown in Figure 5. An icon that is allowed to occupy space can have "parts" that each contribute to the icon's ability to communicate a message to the observer.

« 31 » Redundancy is the key to error correction, and hence to increasing the probability that the messages about fitness encoded by, for example, "4" and "27" can be distinguished. Merely repeating a symbol provides the simplest form of redundancy; for example, the code "11" reinforces the message "1." Three repeats have long been known to be better than two, as in the long-standing Morse-code emergency distress signal:

• • • – – – • • •

or "SOS," by convention always repeated three times.

« 32 » To examine the use of redundancy, we first consider the simplest case, a binary code. For a binary code, the Hamming distance provides a convenient measure of the dissimilarity or distance between two encoded symbols. The codes "111" for "*S*" and "000" for "*O*" are, for example, separated by a Hamming distance of three; three bit flips are required to transform one message into the other. The redundancy of such a code provides a natural sense of spatial dimensionality, as shown in Figure 6. Here flipping a bit is "traveling" in a "direction" on a graph. The bits are independent, so the directions are orthogonal.

« 33 » As can be seen in Figure 6, a three-bit binary code provides the possibility of error correction – every message with mixed bits has a 67% likelihood of being one pure-bit message and only a 33% likelihood of being the other – while the two-bit code does not. Hence a three-fold redundancy is the minimum for error-correction utility for a binary code.

« 34 » At the very basis of human perception is a binary question: is something there or not? It is this question that distinguishes an "object" from an undifferentiated "background." We suggest that the need to answer this simple binary question accurately requires the error-correction capability of a triply redundant encoding and hence a three-dimensional Hamming space. Systems that must answer more complex questions can be expected to employ greater redundancy. This added redundancy comes, however, at a cost: redundant encodings require more degrees of freedom and hence a higher d_{boundary}. Distinguishing the values of these additional degrees of freedom requires, moreover, an energy expenditure of at least $N \times \ln 2 \times kT$ per distinction, where N is the number of bits required to encode each distinguishable value, k is Boltzmann's constant and T is absolute temperature (Landauer 1961, 1999; Bennett 2003).

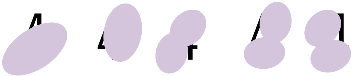


Figure 5 • Spatially encoding an icon allows its "parts" to each contribute to its message.

« 35 » Organisms such as humans do not encode one-to-one eigenform-to-eigenbehavior relationships: there are many different uses for a screwdriver or a coffee cup, and one can reach for and grasp many different objects. We suggest that organisms faced with the task of encoding such complex relationships devote some of their available interface redundancy to encoding eigenform *persistence over time* and the rest to encoding eigenform *actionability*. For example, some degrees of freedom are devoted to encoding that a coffee cup is present, while others are devoted to encoding whether and how it can be grasped. Encodings of persistence and actionability are subject to different constraints. An action type, like grasping, may be executed in a large number of ways, only one of which may yield positive fitness (getting one's coffee!) in a particular situation. Accurately selecting the one right high-fitness grasp from the large number of possible grasps requires a redundant encoding, but redundantly encoding many distinct grasps is expensive. One might expect, therefore, organisms to employ the minimal redundancy that provides error correction, three-fold redundancy, for action encoding. Assuming a continuous range of grasps, a three-fold redundant encoding is an encoding into real ordered triples and hence into real three-space. Discretizing the possible grasps voxelates this space.

« 36 » Employing a distinct real or even a high-resolution discrete three-space for each of a large number of action types would, however, be very expensive both for encoding perception and for memory; one would therefore expect organisms to overlay their encodings so as to encode many different action types in the same space. Whether this is possible depends on the composability of actions and the existence of inverse actions, i.e., on whether the action space supports a group structure. It has been shown, within the CA framework, that a group structure on the action space *G* induces one on the interface *X* (Hoffman, Singh & Prakash 2015). Hence it is plausible to suggest that three-fold encoding redundancy and a group structure on actions is sufficient to generate an interface with three extended "spatial" dimensions in which actions are represented.

« 37 » The encoding of eigenform persistence, on the other hand, is subject only to the constraint of being "good enough" to support appropriate actions. One can, therefore, expect a quasi-hierarchical encoding in which resolution can be varied to suit observational context. As this encoding must "fit into" a spatially-organized interface, one expects a spatial encoding in which the spatial dimensions associated with a particular eigenform are not extended over the entire interface but are rather "compressed" into only a small part of the interface. A compressed spatial structure is a *shape*, like "4" in Figure 5, that occupies space and redundantly encodes persistence.

« 38 » Mammalian visual (e.g., Goodale & Milner 1992) and auditory (e.g., Hickok & Poeppel 2007) systems use distinct processing streams for action and object perception, consistent with the prediction above. Objects are indeed categorized quasi-hierarchically (e.g., Martin 2007). The shapes of both natural and artificial objects can often be represented by scalable codes such as crystal structures, Fibonacci numbers or fractals (e.g., Thompson 1945; Mandelbrot 1982). The idea that spacetime itself is emergent from underlying quantum- or information-theoretic constraints is now being taken seriously by physicists (e.g., Swingle 2012; Arkani-Hamed & Trnka 2014; Pastawski et al. 2015; D'Ariano & Perinotti 2017).

Conclusion

« 39 » In his paper introducing the "it from bit" concept, John Wheeler insisted that "what we call existence is an information-theoretic entity" (Wheeler 1990: 8), later quoting Gottfried Leibniz, "time and space are not things, but orders of things" and Einstein, "time and space are modes by which we think, and not conditions in which we live" in support of his "Fourth No: no space, no time" (ibid: 10). Von Foerster could well have added: spacetime is the eigenform that by remaining constant enables actions.

« 40 » To this we have added: eigenform – eigenbehavior loops, and hence the interfaces through which they pass, encode information about fitness and hence persistence. Spacetime itself, therefore, is an encoding of fitness; it exists only because it is useful to organisms going about the business of staying alive. Organisms with different structures and lifestyles – as different as *E. coli*, an oak tree, and a person – may experience very different "spacetimes."

« 41 » It remains, however, to extract from this idea predictions of sufficient power and precision that confirming them would overcome the intuitive appeal of an "objective" spacetime filled with "objective" objects. The stubborn resistance of the classical worldview in the face of eight decades of quantum theory, experiments and technology shows that this will not be easy. Bringing these ideas into the science – and hence the technology – of perception itself may yet, however, open the door to empirical demonstrations that cannot be denied.

Acknowledgements

We thank Federico Faggin and Manish Singh for discussions and the three anonymous reviewers for their comments. C. F., D. D. H. and C. P. thank the Federico and Elvia Faggin Foundation for financial support.

Received: 20 November 2016
Accepted: 7 April 2017

CHRIS FIELDS

PhD Philosophy, University of Colorado 1985, is an information scientist interested in the physics, developmental biology and cognitive neuroscience of object perception and object re-identification over time. His recent publications include work in the foundations of quantum theory, endophysics, morphogenesis, cognitive modeling, and the etiology of autism spectrum disorders.

DONALD D. HOFFMAN

PhD Computational Psychology, MIT 1983, is a cognitive scientist and author of more than 100 scientific papers and three books, including *Visual Intelligence: How We Create What We See* (2000). He joined the faculty of UC Irvine in 1983, where he is now a full professor in the departments of cognitive science, computer science and philosophy. He received a Distinguished Scientific Award of the American Psychological Association for early career research into visual perception, the Rustum Roy Award of the Chopra Foundation, and the Troland Research Award of the US National Academy of Sciences. Hoffman's research has led to a "user interface" theory of perception, which proposes that natural selection shapes our perceptions not to report truth but simply to guide adaptive behavior.

CHETAN PRAKASH

PhD Mathematical Physics, Cornell University 1982, has published, with Bruce Bennett and Donald Hoffman, the book *Observer Theory*. His current research intends to elaborate a theory that shows how consciousness gives rise to the "physical" world as our interface with reality – as against the idea that brains produce consciousness. As this "reverse hard problem of consciousness" is a view by no means standard in the scientific community, he has used rigorous mathematical analyses to demonstrate the falsity of the commonly held belief that evolution has led us to perceive an "objective" reality with ever-increasing accuracy.

ROBERT PRENTNER

PhD Physical Chemistry, ETH Zürich 2013, has been a visiting scholar at Stanford University's Center for the Explanation of Consciousness. Since Fall 2013 he has been working at the Department of Humanities, Social and Political Sciences at ETH Zürich continuing his philosophical studies and lecturing in the philosophy of science. He is member of the editorial office of the journal *Mind and Matter*.

Open Peer Commentaries

on Chris Fields et al.'s "Eigenforms, Interfaces and Holographic Encoding"

Do Nonclassical Worlds Entail Dualism?

Eric Dietrich
Binghamton University, USA
dietrich/at/binghamton.edu

> Upshot • The vast differences between the objective, classical realm of our everyday lives and any nonclassical realm (like quantum physics) have worried researchers for almost a century. No attempt at resolving the differences or explaining them away has ever worked. Maybe there are two realms, the classical and the nonclassical, and maybe they are paradoxical.

« 1 » Chris Fields et al. are wrestling with, among other things, the paradox, the clash, between "quantum reality" and "classical reality" concerning tables and chairs and dogs and cats and people. There are usually two main ways to deal with paradox. One can try to explain it away (the paradox is illusory) or one can try to eliminate it by showing that one side of the paradox is based on a mistake. Optical illusions are one example of the former way; Zeno's Paradoxes of Motion are an example of the latter way. Of course, there are other, less common ways of dealing with paradox. One can just stipulate the paradox away. This is the method used by mathematicians when dealing with the paradoxes of set theory; this method really only works if one is prepared to go axiomatic. And lastly, one can just embrace the paradox. This is the way taken by paraconsistent logicians, especially those who embrace dialetheism, the thesis that some contradictions are true, while also being false. So, for example, to a dialetheist, the Liar Paradox – "This sentence is false" – is both true and false at the same time. In this commentary, I argue that though the authors opt for an eliminativist approach to the nonclassical-classical paradox, they ought to opt for the last way: they ought to embrace the dualistic paradox.

« 2 » In their article, Fields et al. present an interesting and large theory that begins with taking observer-relativity seriously and ends with the proposal that spacetime could profitably be construed as error-correcting code. Then at the end, in §41, the authors say that their theory still needs to produce predictions sufficiently powerful to overcome the intuitive appeal of mind-independent spacetime filled with mind-independent objects – i.e., powerful enough to overcome our resolutely perceiving the classical world.

« 3 » In the very next sentence, the reader senses perhaps some despair on the part of the authors, for they bemoan the "stubborn resistance" of the classical *world* in the face of eight decades of quantum theory – in effect saying that after eight decades, one would have thought that we would have finally said goodbye to the classical world, to the mind-independent world. Interestingly, perhaps in an effort to hurry the classical world out the door, the authors do not use the term "world," but rather call it a *worldview*. But this latter is a term they are not entitled to because, as they just said, they have yet to prove their theory experimentally because they have yet to derive any experiments from their theory. For all they know now, it seems, the classical world *is* the world, or at least one of them. There are not merely different viewpoints, rather there are different worlds.

« 4 » The authors, then, are stuck with the classical, mind-independent world while they develop and experimentally test their new theory, which posits a nonclassical, mind-dependent "world" as a replacement.

« 5 » It is not clear what the authors hope for at this stage. They themselves are acutely sensitive to the staying power, the stubbornness, of the classical world. But they also know the explanatory power of mind-dependent approaches to understanding minds and their realities (there are many reasons to take observer-relativity seriously). One gets the impression that by drawing from several sources – quantum physics, consciousness studies, cognitive science, evolutionary theory, math, and philosophy – the authors hope that their theory will simply liberate the human mind from its preference for occupying a mind-independent universe.

« 6 » At this point a movie reference is needed. In the movie *Arrival*, space aliens show up in the present time and offer us the gift of their written language. This language is unlike any language on Earth. To use it, one has to have a decidedly nonhuman relation to time – in particular, one has to be able to see the future. To the space aliens, seeing the future is second nature; indeed, they experience all at once what we would call sequential events. The key is that when humans learn the alien language, their perception of time changes, and, like the space aliens, they then also see the future, experiencing all events at once. Learning their language changes our brains. Do the authors want the same property for their theory – merely learning it, or learning that experiments support it, will change our stubborn human resistance to sensing the world in a mind-dependent fashion? Will learning their theory, or learning that their theory agrees with all experimental challenges, change human perception in such a way that the classical world is eliminated?

« 7 » Of course, it is unlikely the authors want any such thing (still, in §41, they do say that confirmatory predictions of their theory would "overcome the intuitive appeal of [the classical world …]"). Assuming the authors do not think mere knowledge of their theory will liberate humans from our classical world or diminish its appeal (this has not worked for quantum mechanics), then what are they going to do about "the stubborn resistance of the classical world"? Unless something frees human perception from its moorings in the classical world, it does not matter what brave new theory is developed, the moorings will remain.

« 8 » Suppose *X* is the extremely sophisticated future version of the theory the authors are working on now, and thus considered "ultimately true." It is profoundly unlikely that *X* will finally free humans from their classical worldview (the movie, after all, was fiction). Rather, we will be stuck with the very situation we have now with quantum physics, where human physicists occupy the classical world while they develop, experiment on, and prove the nonclassical theory of quantum reality. We have had 80 years of quantum mechanics (as Fields et al. note). In that whole time, no physicist has started experiencing the nonclassical world in their daily lives. Rather, they all daily experience the classical world. And these physicists also experience the classical world while they experiment on and theorize about the nonclassical world. So, the authors' theory, *X*, will represent a nonclassical realm, and we will learn it, apply it, and come to see *X*'s beauty, all the while firmly planted in the classical world. Go back and watch the videos of the announcement at the Large Hadron Collider (LHC) of finally finding the Higgs Boson. Everything in the video is classical. The Higgs is not. The same can be said of X.

« 9 » So, what to do? We humans seem to occupy one realm, the classical one, while developing nonclassical theories of nonclassical realms accessible to us only via our thought (the LHC is classical, the data from its experiments are classically presented and represented, but via our minds, we see beyond the data to a nonclassical world). And the two realms together form a paradox: crucial propositions true in one realm are false in the other.

« 10 » One proposal is to give up the quest to "overcome the intuitive appeal" of the classical world (§41). Embrace the two worlds, or many worlds, solution: one is classical and others are not.

« 11 » Specifically, the authors' theory could explain human and other animal minds in the nonclassical way they detail, while at the same time, we humans and other cognizers occupy a classical world.

« 12 » I said above (in §3) that the authors were not entitled to use the term "classical worldview" (from their §41) because until their theory was supported by experiments, they could not know that classicality was a world*view* and not a *world*. We now see that "classical worldview" has another problem. It suggests that there is one world: from one worldview (point of view) it looks classical and from another it looks nonclassical. Think about walking around a car. From one view (a sideview) the car looks one way, from another (a front view) it looks another. The "real" car is the integration of all such views (for the viewer). Note that the car is *not* paradoxical, so the integration works. But this does not apply to the world posited by the authors' and the classical view we inhabit as we read about their theory: the two are decidedly paradoxical. So, integration is unlikely to work. The one-world-with-two-worldviews approach might, I suppose, better accord with Ockham's Razor, but that's not in the cards. This all suggests that there are many *worlds* – we view them somehow by visiting them, by "changing locations," via our consciousness. (Of course, ontologically, some of us are still committed to some over-arching, single *meta-world*, and this meta-world has to be at least contradictory and probably dialetheic (the locus of unresolvable contradictions). As with the other issues in this area, it is not clear why this meta-world appears or exists. I am inclined to invoke the observer, which is what I think the authors might support.)

« 13 » The cost associated with this contradictory-worlds approach, and not just contradictory points of view, is that *consciousness* remains unexplainable. But many of us already think this is the ultimate knowledge about consciousness (Dietrich & Hardcastle 2005). It is unlikely that the authors will agree with this since a large part of the motivation for their article is bringing consciousness into the science tent.

« 14 » Regardless of whether one picks one world with many contradictory, paradoxical viewpoints or many contradictory worlds, the (unintended) message of the authors' research seems clear: the classical world does not merely have an "intuitive appeal" for us (§41), rather it is *ineluctable*. We are classical beings with minds that allow us to see the nonclassical. How this can be so is very puzzling. And the authors' theory does not directly address this. However, as already claimed in §8 above, it is very unlikely that any theory of this "dualism" – classical beings studying a nonclassical realm – will ever be intuitive to us even though it may well be robustly explanatory. What will come to seem intuitive then is that what is called "reality" is bigger than we thought, and more unstable and protean than we supposed. Epistemic humility should follow.

Eric Dietrich is professor of philosophy at Binghamton University in New York and the editor of the *Journal of Experimental and Theoretical Artificial Intelligence*. He is the author of the book *Excellent Beauty: The Naturalness of Religion and the Unnaturalness of the World* (2015) and the paper "There is no progress in philosophy" (2011).

Received: 31 May 2017
Accepted: 8 June 2017

Where is Spacetime Constituted?

Urban Kordeš
University of Ljubljana, Slovenia
urban.kordes/at/pef.uni-lj.si

> Upshot • In an attempt to understand its presuppositions, the commentary takes a closer look at the model proposed by the target article. By analysing the interactions between conscious agents, the model tries to derive the enaction of a spacetime framework. A critical examination of the ontological status of the involved entities indicates inconsistencies, especially at the adoption of viewpoints. It seems that despite the model's being supposedly grounded on the primacy of consciousness, this characteristic is not immediately apparent. The commentary proposes an even more radical adoption of the first-person point of view.

Ontological status of entities in the conscious agents model

« 1 » I am inclined to support the model presented by Chris Fields et al., especially the way it, in one big stroke, connects biological constructivism (Maturana & Varela 1980; Foerster 1984; Riegler 2012) with quantum physics. Yet, extraordinary claims (such as the proposed model) require extraordinary evidence. When the model's results confirm the authors' goals, i.e., that from the interactions of conscious agents almost miraculously springs a 3 + 1D physical framework of our everyday world, one should always beware of the possibility of motivated reasoning.

« 2 » As the remainder of this section will show, an explication of the proposed model's presuppositions exposes considerable issues. It remains to be seen whether those problems stem from the commentator's misunderstanding, from small inconsistencies in the proposed model (which can be easily patched), or from flaws with serious consequences for the model's fitness. I hope it will turn out to be one of the former options, for the idea of deriving characteristics of the physical world from the dynamic of consciousness is an exceptional one.

« 3 » The aim of the target article is to create a mathematical model of how consciousness constitutes the world. The authors avoid the presupposition of "objects as spatially-bounded, temporally-persistent, internally-cohesive, causally-independent entities" (§14), and instead attempt to create a mathematical model of the constitution of those objects, presuming the primacy of consciousness. Discarding the natural attitude (the tendency to believe our construction of the world to be an accurate representation of objective reality), the authors seem to assume the phenomenological attitude (Husserl 1982), the attitude that phenomenology shares with constructivism (as argued in Kordeš 2016a).

« 4 » According to phenomenology, phenomenal consciousness is the epistemologically safest foundation on which to build science. According to Dan Zahavi (2004), for Edmund Husserl, studying how the world is constituted in consciousness became the cornerstone for transcendental phenomenology, which in turn was supposed to become the foundation of science. Despite the fact that Husserl created a philosophical system with this particular purpose, phenomenology has never completely succeeded in this endeavour. The problem being that phenomenologists never made it exactly clear how to actually build natural science (starting with physics) on phenomenological foundations. The target article offers a solution.

« 5 » The proposed mathematical model is based on the concept of conscious agents (CAs) (§2). In the following paragraphs I will try to summarise and more clearly explicate the presuppositions that come with this concept.

« 6 » The authors suggest that a defining feature of a CA is its "principled unpredictability […] considered by some to indicate autonomy or 'free will' and hence agency from the perspective of external observers" (§6). Furthermore:

> "While autonomy in the non-trivial machine sense inferred above is somewhat abstract, a requirement for autonomous decision-making at least suggests an awareness of potential consequences and hence consciousness." (§8)

« 7 » From this definition of a CA, it is clear that consciousness is inferred from the CA's behaviour. Since this behaviour takes place in an abstract mathematical space rather than in the space of our everyday world, the question arises: What is the ontological status of entities or phenomena those spaces represent?

The gap between functional and phenomenal aspects of consciousness

« 8 » Susan Blackmore (2013) divides discussions concerning consciousness into two distinct realms represented by the following two questions: "What is it like to be…?" and "What does consciousness do?" (for the purposes of this commentary, they will be referred to as the phenomenal and the functional aspect respectively). There are many answers to the latter. One of them is proposed by the target article, i.e., consciousness behaves in principle unpredictably. Between the functional aspect of consciousness and the aspect that answers the question "What is it like to be…?" (describing so-called phenomenal consciousness), there is an unsurmountable chasm – usually referred to as the explanatory gap.

« 9 » In order to assess which aspect is assumed by the authors of the target article, the basic mathematical elements of the proposed model need to be examined. What are the categories that define agents CA_1 and CA_2, the interaction between whom enacts physical entities? Figure 1 of the target article provides the answer: "Here X_1 and G_1 and X_2 and G_2 are measurable spaces representing the experiences and available actions, respectively." The space X is especially important as on it rests the weight of the entire model. It is precisely X that is supposed to contain encoded objects.

« 10 » But what kind of entities does X represent? What is the meaning of "experiences" (§9) within the model? It would seem that X also introduces phenomenal consciousness into the model based on the strong presupposition that phenomenal consciousness can be mathematically described. With this, the model adopts the first-person perspective of lived experience (a perspective that is unreachable for most of natural science). By simultaneously including the functional and the phenomenal aspect of consciousness it seems that the model of Fields et al. unwittingly mixes first- and third-person perspectives.

« 11 » Another indication for the mix-up of perspectives is the model's separation of *G* and X. Separating experiences (*X*) and available actions (*G*) indicates a distinction between the two. If the model took the autonomy of CAs and the primacy of experience seriously, *G* would be a subset of *X* – available actions are only those noticed or autonomously constructed and as such experienced as available by the CA. Because that is not the case, the only possible interpretation is that the authors presuppose the possibility of a space of available actions as perceived from outside the CA. This takes autonomy away from the agent. Being autonomous means that the agent chooses from the options the agent itself constructs rather than from pre-given options (cf. Winograd & Flores 1986). Genuine autonomy is in the very construction of the elements of the world, which are, in this case, options to choose from.

Consciousness as the foundation

« 12 » With the exception of phenomenology, most other approaches see consciousness as a product of an observer-independent, "natural" world (i.e., they naturalise consciousness). If consciousness is to be taken as the foundation of a theory, then naturalising approaches are inappropriate, as they presuppose the primacy of something other than consciousness. The only aspect of consciousness that can be used as the foundation for a theory is phenomenal consciousness, i.e., lived experience. This is only possible if the theory's point of view is a first-person one. However, in the case of the proposed model it is the point of view of the CA.

« 13 » Constructivists always stress that every view is a view from somewhere. I fear that Fields et al. are not very clear from where they are observing. Are they looking at the world from the eyes of an agent (who, of course, does not have access to anything other than its own horizon – i.e., the surface that connects it to the world) or through the "eyes of God," who sees all agents, their actions and interactions?

« 14 » The "God's eye" view or the view "from nowhere" (Nagel 1989) is characteristic of fields that have uncritically accepted the natural attitude (that is, for most of science with a few exceptions, such as phenomenologically inspired research). This view enables intersubjectively valid methods and exceptionally successful research, characteristic of physics, neuroscience, biology, etc. What this view filters out, though, is consciousness. It perceives the researched structures as "real" and forgets that they came about only due to the act of consciousness. If naturalising research approaches are at all interested in consciousness, they look for it as a product of those natural structures. By filtering out the observer's consciousness, the naturalistic view can only resort to inference from behavior when trying to detect consciousness "out there." As a consequence, they can only answer the functional question, i.e., "What does consciousness do?" while the question of phenomenal consciousness – "What is it like to be…?" – is inaccessible to the behaviour-oriented third-person view of natural science.

« 15 » By renouncing the view from nowhere, consciousness appears everywhere. Phenomenal consciousness imbues everything there is, everything one notices, thinks or perceives (Kordeš 2016b). Consciousness from the first-person perspective is a medium in which all features of the world are constituted.

« 16 » The history of cognitive science has shown that the growing understanding of brain dynamics and human behaviour does not bring us closer to understanding experience. The failure to bridge the explanatory gap points towards the conclusion that phenomenal consciousness is not only primary but also irreducible. If we want to get conscious experiences as a result, we have to start with conscious experiences. Only in that case can we say that we take consciousness as the foundation of our theory.

« 17 » The model proposed in the target article puts agents and their life dramas in an abstract space. The authors attempt to "develop the dynamics of interacting conscious agents, and study how the perception of objects and spacetime can emerge from such dynamics" (Hoffman & Prakash 2014: 557). Whatever this space is supposed to represent does not seem to represent the space of phenomenal consciousness. As argued above, only if the theory performs the (very radical) step of grounding itself in phenomenal consciousness, is it sensible to start looking for appropriate mathematics that might enable the modelling of the constitution of the world. (One of such notable attempts being "primary algebra" proposed by George Spencer Brown 1969 in his *Laws of Form*).

« 18 » It would seem that the authors are not modelling the construction of a world from consciousness, but the construction of a world by entities that are behaving as if conscious.

Agency and the sense of agency

« 19 » The confusion of perspectives is also apparent from the use of the term "agent" and the consequential notion of agency. It seems that the authors conflate the sense of agency with agency as the actual ability of a CA to consciously influence courses of action. Agency and the sense of agency should not be carelessly equated. Many third-person studies such as those of Benjamin Libet et al. (1983) and Daniel Wegner (2003) have shown that our conscious decisions are not (always) causally linked with our actions, despite what the sense of agency might suggest. The phenomenal sense of agency functions mostly as a way of smoothing the narrative (i.e., sense-making).

« 20 » Agency and the sense of agency could only be equated if the model were to be intrinsically rooted in the experiential world, that is, if the whole process were to be seen as metamorphoses of phenomenal consciousness. Such a model would describe a consciousness that changes itself. That way, sense-making, the constitution of objects, etc. would all be part of the same substance, and the dualism that spoils the image of the presented model would be avoided.

Acknowledgement

I would like to thank Florian Klauser for his help with exploring and articulating the target article's (as well as our own) epistemological positions and presuppositions.

Urban Kordeš is professor of cognitive science and first-person research at the University of Ljubljana where he is currently heading the cognitive science programme. His current research involves such training in order to study the phenomenology of the enactment of knowledge.

Received: 6 June 2017
Accepted: 14 June 2017

Eigenform Encoding and Spacetime

Louis H. Kauffman
University of Illinois at Chicago, USA
kauffman/at/uic.edu

> Upshot • An eigenform is both a symbol for a process and the instantiation of a process itself. As such, eigenform provides a new entry to spacetime, as a unification of entity, place and process.

What is an eigenform?

« 1 » In order to provide some background for a discussion of the target article "Eigenforms, Interfaces and Holographic Encoding" by Chris Fields, Donald Hoffman, Chetan Prakash and Robert Prentner, I shall start this discussion by describing what an eigenform is and then I shall explore the nature of the relationship between quantum theory and eigenforms. First, let us note that formally, mathematically, an eigenform is nothing more and nothing less than the fixed point of a transformation in some domain. If the domain has name D and the transformation is regarded as a function T: $D \rightarrow D$, then an eigenform E is an entity (either in D or in an extension of D) such that $T(E) = E$.

« 2 » Why do we take this notion of eigenform to be of importance for cybernetics? An initial answer is that the transformation T acting on a system D produces a natural recursion. Start with $X(0)$, some entity that we think may approximate a fixed point. Let $X(1) = T(X(0))$. In general, let $X(n+1) = T(X(n))$ for $n = 1, 2, 3, \dots$ ad infinitum. Then the transformation T becomes the generator of a process and hence propels the system into time by the very action of the transformation. This process may have no fixed point. And we are well familiar with such a situation. In fact, almost every object or action that we know has a potentially endless recursion associated with it. This applies in particular to fundamental transformations, such as simple motions of the human body like taking an upright step. We take a step and we can take another. Of course some transformations do have fixed points. For example, $T(x) = x^2$ has as a fixed point the number 1, whose square is equal to itself. Alas, this fixed point will not be reached if we take a starting value that is not equal to 1. If we start with a number greater than 1 and square it, we get a number even greater than that and the values will approach infinity. Infinity! Well we were not thinking of that as a number, but surely Infinity^2 = Infinity and so Infinity is (if we allow it into our conversation) an eigenform for T. If we take a number greater than 0 and less than 1, then applying T to that number will lead to a sequence that tends to 0. And 0 is a fixed point of T, indeed. So, we have found that T has three eigenforms, Infinity, 1 and 0. This could lead us out beyond the specific transformation to thoughts about the fantastic distinction that seems to present itself between the Infinite, the Nothingness and Unity. We could go off track as far as the calculating forms are concerned and find that the simple working with and searching for a fixed point for $T(x) = x^2$ has led us into cosmological concerns.

« 3 » Heinz von Foerster, in discussing what he called "eigenvalues" (Foerster 1981) and what I call "eigenforms" went off track in a carefully planned formal way that indicates a systematic abduction from the given system into a larger context. He suggested considering the context-free application of T upon itself, for any T whatsoever! And he finds that he can take $E = T(T(T(T(T(\dots)))))$ and then with this infinite concatenation of T upon itself, like the deep repeated reflections seen by an observer between two mirrors, we have $T(E) = E$. What has happened here? Does this concept go too far? Any T has a fixed point and that fixed point is nothing more than an infinite reflection zone of copies of T in a circuit upon themselves. Such a fixed point has no basis other than the transformation T itself. John Wheeler (Misner, Thorne & Wheeler 1973) had the same concept for quantum cosmology. He said (in my paraphrase) that *the Universe is a self-excited circuit, arising from its own observation of itself, which is that very observation of itself.* There is nothing in the universe except the self-participation of the nothing that becomes information and form arising from its own eternal return. The eigenform E is an existence and comes about in the cleft where spatial form and temporal process (time itself) meet. Von Foerster pronounced this self-excited circuit in his own way with his statement "I am the observed relation between myself and observing myself" (Foerster 1981). We can go from von Foerster to Wheeler by a substitution: "The Universe is the observed relation between itself and observing itself." There is no difference. Spacetime, the Universe, the Self, all are central eigenforms in the genesis of worlds. These words are here capitalized to indicate their roles in this allegory of the nature of Everything.

Quantum theory and it from qubit

« 4 » Having stated my point of view, directly and allegorically, let us turn to the target article, where the authors say "[…] we pursue the notion of an eigenform not from the point of view of an abstract reflexive system, but from von Foerster's original perspective of an agent that observes and acts on its world" (§2). This is a correct stance. One can consider an abstract reflexive system, but the whole point in considering a reflexive system is that the agent, the observer, is the system, and observers become both the system and the parts of the system. Let the allegory become prose. The universe is the source of its own observation. The universe is a self-excited circuit. The agents are not separate from their worlds. In §2, Fields et al. say that we propose an "epistemic cut" between agent and world for the purpose of theory construction. Theory demands such a cut in order to distinguish a theorizing agent. In fact, such a cut has to come along with any perception at all. And the key to the situation of perception is that we are sensitive to the fact that while a distinction is made, it is also mutable. There is no final cut and in the acts of perception, as we come to our senses, we find those places of ambiguity, of feeling, where it is not possible to say what is our construction and what is the world.

« 5 » In §3, the authors state:

> "We suggest that spacetime itself, including both the space in which objects appear to be embedded and the time over which they appear to persist, is a relational, error-correcting code for the fitness consequences of interactions."

At this point I am not prepared to comment on the nature of the code as error-correcting. I am not clear what constitutes an error in the context of the authors' position. Thus, I ask the authors for clarification about their notion of error (Q1). At one level there can be no error. At another level, what are called errors are certain distinctions made by an observer. At yet another level, errors are what are "corrected" by feedback. In the case of eigenform, there can be "error correction" in the sense of stabilization if the recursive process does stabilize. But we also create forms of stabilization such as the infinite concatenation of the agent's action in the form of $E = T(T(T(\ldots)))$. It is to be understood that this infinite activity is an *abduction*, a leap to a form that is invariant under T. It does not mean that an infinite number of operations have been performed. It means that the self-excited living circuit has come into being. In this sense, the eigenform corrects itself. It *makes* itself correct. I am the one who says I.

« 6 » As in §12, the authors are quite taken with the notion of physical surface as the manifestation of the epistemic boundary. I agree that this is useful for physics and particularly in the wake of the recent holographic hypotheses in cosmology and quantum physics. But the most generally applicable epistemic boundary is any distinction whatsoever. And when I say distinction, I mean an arising of observed difference and the arising of an observer of this apparent difference. *I do not regard physical surfaces as fundamental epistemic distinctions. I regard them as eigenforms for the convenience of the observer who is searching for deeper understanding.* Can the authors address this issue from their point of view? (Q2)

« 7 » To offer support for my point in §6, above, consider the fundamental situation of the quantum mechanical model. The state of a quantum system is a vector in a complex vector space (a Hilbert space) that is seen as a sum over all possible observations that can be made for the given experiment. These possible observations are taken as an orthonormal basis for the vector space, and the sum of the absolute values squared of the coefficients of the basis vectors is equal to unity (now lower case). In this way, the state vector is a probability distribution and indeed the probability of making an observation of one of these possibilities is equal to the absolute square of its coefficient. Physical processes are unitary transformations that preserve the total probability distribution. Because we allow complex coefficients, the superposition can model interference and quantum effects. An observation makes the distinction that brings forth one of the possibilities. This distinction is often articulated without the usual spatial boundaries. Thus, a superposition that indicates an entangled state does not have to show the spatial structure that may possibly separate the entangled particles. We need only know the form of their entanglement to know that upon observing one of them, the other's possible observation is determined. This interrelationship goes across the structure of spacetime. This situation is seen by some to be a paradox. I state it here to bring into question the notion of Fields et al. that the fundamental source of the epistemic boundary is spacetime itself. The Universe (now captitalized as we reenter the allegory) goes beyond spacetime and comes forth as self-excited circuit, living quantum information, unified with a living observer that is both distinct and not distinct from what is observed.

« 8 » I believe that it is a fundamental insight that Universe is identical with Self and that it shall be possible to reformulate present-day physics so that it is seen as a form of the living. We are not there yet. Fields et al. have gone forward with courage to explore aspects of this possibility for Unity.

Louis H. Kauffman is Professor of Mathematics at the University of Illinois at Chicago. He has a BS from MIT and a PhD in Mathematics from Princeton University. He has been President of the American Society for Cybernetics and is a Fellow of the American Mathematical Society and a recipient of the Warren McCulloch and Norbert Wiener medal of the American Society for Cybernetics. His research is in cybernetics, form, topology, recursions and the theory of knots.

Received: 7 June 2017
Accepted: 14 June 2017

Certain Questions Regarding Perception and Boundaries

Konrad Werner
University of Warsaw, Poland
konrad.t.werner/at/gmail.com

> Upshot • I elaborate on how boundaries are accounted for in the target article. This is a substantial issue if we are to understand the proposal laid out by Fields et al. I argue that certain boundary-related notions and theses need clarification.

« 1 » Perception has to do with boundaries. I find this general idea laid out in the target article by Chris Fields et al. very intriguing. The text is insightful in many respects, yet it is also dense, which makes addressing all the issues that should be addressed virtually impossible. For that reason, I shall focus solely on the problem of boundaries. They are salient factors in the proposal under consideration, however, certain things need clarification. There is a literature on boundaries, in particular, a sub-discipline of ontology called mereotopology (see references in my target article in this issue). I shall not refer to this literature, though, as it would require much more space.

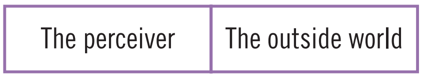
« 2 » So, what does perception have to do with boundaries? The authors say that the perceived world, the familiar realm of things that we perceive every day, results from encoding the incoming data for us and that this happens on the boundary that separates us from our surroundings. However, crucially, the perceived spectacle does not encode "information about the ontological or causal structure of the world, but rather information about the structure of the fitness function that relates the agent to the world" (§3). This means that the outside world is a black box (a metaphor brought up several times by the authors themselves): that it is, in a sense, hidden.

How many boundaries?

« 3 » Imagine that you are on a beach and you see the line dividing the surface of the sea and the sky above it. Now, how many boundaries are there and to which entity do they belong? Is it the case that the water (or the air) is *closed*, meaning that it has its boundary as a part, while the air (respectively, the water) is *open*, i.e., it does not have a boundary of its own, thus the boundary of one entity serves as the boundary between the entities in question? Or is it the case that both the water and the air have their own boundaries and that these boundaries abut each other? Finally, perhaps the water and the air share the boundary, meaning that the latter is a common part of both entities. (I omit the antirealist scenario in which there is no boundary at all but only an illusion of its actual existence). These questions may seem silly, yet if being bounded is an essential feature of some entity, a condition of its identity, the issue becomes ontologically critical.

« 4 » When it comes to our case, the authors claim that perception is a spectacle played on the boundary between the perceiver and the outside world (the black box), but how many boundaries do we have there? Here is the first option:

(1B*pw*) There is one boundary. It belongs to the perceiver and to the outside world; it is shared by them. They are both closed.



This option is clearly endorsed by the authors in §10.

« 5 » If we generalize what the authors say in §10, then the perceiver and the outside world are *neither closed nor open* (originally, the model outlined in the target article referred to the simplified situation in which an outside world for one perceiver is another perceiver). This cannot be right. The alleged "purely notional" character of the boundary in question has nothing to do with the context. This is because, by deeming the boundary "purely notional" we take a particular position as regards the *nature* of the boundary, not about its very *existence*. So, for example, the boundary between Poland and Russia (Kaliningrad Oblast) is purely conventional, but surely it exists and it is even guarded by heavily armed forces. So, if there is a boundary (or boundaries) between the perceiver and the outside world, regardless of its nature, the two realms must be either open or closed.

« 6 » So, suppose that we have one boundary shared by both sides as (1B*pw*) proposes. This means that both the perceiver and the world are closed, yet they are not separated. Imagine two pieces of a material sewn together: they are distinct and each of them is bounded but they cannot be set apart; they are parts of one whole, so to speak, precisely because they are sewn. However, there is one subtle puzzle here: if the perceiver and the world are sewn by their shared boundary, then one can hardly say that what happens in the sewing itself has nothing to do with the ontological structure of the world; after all, this sewing is likely part of the ontological structure; if not, then what is it?

« 7 » In this context, we can notice an interesting tension in the very nature of at least some boundaries. Think of a living creature: boundaries constitute an organism by cutting it off from its environment, yet at the same time, they provide channels for communication with the environment. Say, once they bound something, they open some doors to make traffic possible. When it comes to the philosophy of mind and perception this tension is crucial: there is the Cartesian approach to the mind-world boundary, putting stress on isolation or separation, while, e.g., in Edmund Husserl's or Maurice Merleau-Ponty's approaches, the boundary in question was supposed to – let me use Husserl's original and very pregnant formulation – *bring* the world to *a presentation*. The authors apparently take the Cartesian route and I am not sure if that is necessary for their project as a whole.

« 8 » But perhaps there are actually two boundaries, as it is also suggested in §10, where the authors introduce a distinction between a boundary and its surfaces. But is a surface not a boundary, too? So, we can at least take into consideration the following scenario:

(2B) There is one boundary that belongs to the perceiver and one boundary that belongs to the outside world. They are both closed.

However, this case is very problematic due to the fact that it becomes unclear where exactly the information is encoded: on which boundary does this process occur? If it occurs on the perceiver's boundary, then what role is left to be played by the world's boundary? Perhaps here is the point where the idea of the *structure of fitness*, as opposed to the ontological structure of the world, comes on stage. Suppose that the world's boundary provides a barrier that the perceiver bumps against, so to speak, adjusting its shape, i.e., adjusting its boundaries, so that they fit, metaphorically, to the world's boundaries. However, if there are two separate boundaries and their abutting determines the structure of fitness, then why is there any need for a rather complex process of *encoding* information and establishing this whole theater of phenomena that we face once we open our eyes in the morning? This is just another way of formulating what David Chalmers (1995) once called the hard problem of consciousness, yet from a different side; this is, say, the *hard problem of presentations*: why there presents something rather than nothing; why are we not "zombies," bumping against the boundary of the world, adjusting to it and by doing so maintaining solely our structure of fitness? It seems that we could do so without facing any phenomena and it is likely that the most primitive organisms still function in this way.

« 9 » Perhaps stripping the perceiver from its boundaries yields an even better understanding of the structure of fitness.

(1B*w*) There is one boundary and it belongs to the outside world. The perceiver is open while the world is closed.

Here the perceiver is shaped by the boundaries of the world as boundless water poured into the glass. In this sense, the perceiver fits the boundary (or boundaries) of the world. This boundary must be there, pre-given and ready-made (Hilary Putnam's term) independently of the perceiver if the latter is supposed to adjust itself to it. Such a scenario has been discussed and criticized, e.g., by Francisco Varela, Evan Thompson and Eleanor Rosch (1991: 193, 198). However, aside from Varela's criticism, here as in the (1B*pw*) scenario, it is not clear why the structure of fitness is distinct from the alleged ontological structure of the world. After all, the boundary that the perceiver faces partakes in this ontological structure. If it does not partake in it, then what is it? Thinking of the boundary in question as if it were like a mere wrapping paper having nothing to do with the thing being wrapped – the world in this case – makes the boundary a mysterious, superficial entity of unknown origin. But if it is not a mere wrapping paper, then one cannot say that the ontological structure is hidden behind the boundary; the structure is there, and the boundary is its manifestation.

« 10 » Finally, there is a scenario that strips the world from its boundary:

(1B*p*) There is one boundary and it belongs to the perceiver. The world is open while the perceiver is closed.

Varela et al. likely have this scenario in their minds when they write that “our lived world does not have predefined boundaries” (Varela, Thompson & Rosch 1991: 148), and

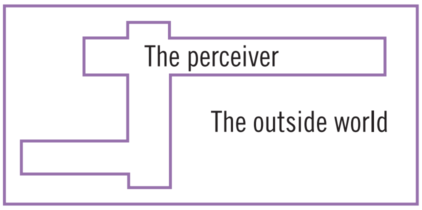
> “cognition is not the representation of a pregiven world by a pregiven mind but is rather the enactment of the world and a mind on the basis of a history of the variety of actions that a being in the world performs.” (ibid: 9).

Here, admittedly, we cannot speak of a pregiven or ready-made ontological structure of the world in the absence of what the perceiver does. However, whatever the structure of this world is, it cannot be regarded as obscured or hidden either. Here, structures of the world result from the perceiver’s interactions with the world.

« 11 » To conclude this part, I wonder to what extent the conception outlined in the target article could be freed from the Cartesian idea of the world’s being ready-made (its having a structure independently of cognition) and hidden (meaning the inaccessibility of this ready-made structure).

How many types of boundaries?

« 12 » The scenarios presented above, as well as those presented in the target article, suggest that there is just one place where the perceiver-world boundary is drawn so that both sides seem to be like two blocks. But what forces us to accept this two-blocks model? Maybe it would be much better to draw several lines composing a more complex structure, say, something like this:



« 13 » Someone might say that there is no essential difference between the latter and former scenarios. While this is correct, the latter drawing makes an important suggestion: both sides, i.e., the perceiver and the world, are *shaped with respect to each other*; the boundary line is not just *a line*; it contributes to what the two bounded realms are.

« 14 » There may be, however, an essential difference, too. The essential, yet rather tacit assumption behind what I have just dubbed provisionally the two-blocks model is that there is just one *type* of perceiver-world boundary. But why? Perhaps each perceptual subsystem, be it vision, hearing or touch, sets up and imposes on the world its own structure of boundaries. Note that from the evolutionary perspective, the step from mere mechanical senses like touch or from chemical senses, the oldest ones, to vision – the step that marks a great evolutionary achievement – originated from a new ability to target what was literally on the boundary of an organism, where receptors are plugged in, not as the object perceived but as a signal of an object or as information. Hence, while in the case of touch, the boundary of the thing being perceived abuts the physical boundary of the perceiver (let alone chemical sensation where a substance that is perceived must react with certain proteins, which makes the question of boundaries difficult – there is something more than abutting), in the case of vision, for instance, these respective boundaries have nothing to do with each other. But perhaps – let me set this off as a speculative hypothesis – together with vision, a specific *new system of boundaries* came into being, so that, say, the vision-determined boundary of the perceiver is not identical to its physical boundary *qua* organism, and at the same time this new boundary serves as the vision-determined boundary of the thing perceived. Here perception, cognition in general, brings forth significantly new types of boundaries and – this is a constructivist aspect of the idea – imposes these boundaries on the world so that the world is brought to a presentation in such and such a guise (see my target article in this issue). And perhaps further steps in this evolutionary process resulted in the boundaries of what we used to call *mind*. Recall Andy Clark and Chalmers’s (1998) groundbreaking idea of an extended mind. What they propose boils down to the claim that the mind sets its special arrangement of boundaries that are not identical to the physical boundaries of the body.

Konrad Werner works at the University of Warsaw, Poland, where he is running a research project in experimental philosophy. He obtained his PhD in 2013 from the Jagiellonian University in Krakow. He is interested, first of all, in so-called “aspectual shape” or perspective-dependency of perception and knowledge. His articles concern epistemological, ontological as well as logical aspects of this issue. Still in connection with it, he has recently directed his attention toward evolutionary origins of cognition.

Received: 19 May 2017
Accepted: 20 June 2017

"Eigenforms, Interfaces and Holographic Encoding": Their Relation to the Information Loss Paradox for Black Holes and Quantum Gravity

Antonino Marcianò
Fudan University, China
marciano/at/fudan.edu.cn

> Upshot • I emphasize possible analogies and links between the content of Fields et al.'s target article and some consolidated recent studies in the literature of quantum gravity and the information loss paradox for black holes. This follows from the attempt by the authors to account for spacetime as an error-correcting code. The paradigm the authors focus on can be naturally cast in the language of some models of quantum gravity based on graph theory, and suggests a generalization of the perceptual systems so as to account for quantum holographic encoding as described in quantum gravity.

« 1 » At the core of the target article "Eigenforms, interfaces and holographic encoding: Toward an evolutionary account of objects and spacetime" there is the development of the interface theory of perception (Hoffman, Singh & Prakash 2015). This framework is unfolded within the very same language in which the epistemic foundations of quantum mechanics can be phrased (§§4–13). The interface theory of perception allows a detailed description of the holographic encoding, and is naturally tailored in order to account for the complexity of the observer–environment interface's interactions. Within this framework the authors address the structure itself of spacetime (§§32–36), after having reviewed and analyzed the most relevant options for holographic encoding (§§15–18), and summarized the propositions for the fitness functions (§§22–24) deployed in the interface theory of perception.

« 2 » Although the axioms of quantum mechanics are not explicitly stated, the focus throughout the work is on quantum states, entanglement and observer-environment decoherence (§§15–18). The underlying assumption is therefore that a theory of perception should be addressed and studied through the lenses of quantum mechanics. The authors explicitly mention criticism of the assumptions of "epistemic transparency and objective persistence" proper of the classical worldview and point toward the elaboration of experiences within the theoretical framework of quantum mechanics. They ascribe particular relevance to unitary quantum theory (§15) as the correct paradigm in which to address decoherence and holographic encoding. As they point out, in unitary quantum theory "the universe is permanently in an entangled state; there *are no* classical objects" (§15; emphasis in the original).

« 3 » Besides the philosophical preference toward unitary quantum mechanics, the line of thought followed by the authors has a striking overlap with a vast part of the literature developed in the last four decades about the information loss paradox of black holes, and crosses its natural consequence, which is the development of the holographic principle – see, e.g., the seminal works by Gerard 't Hooft (1993) and Leonard Susskind (1995) – in quantum gravity and high energy physics (Bousso 2002). There are evident analogies that assimilate the crucial role of the black hole event horizon in the flow of information to the role of the membrane between observer and system in the interface theory of perception. The comprehension of the function of the physical degrees of freedom that puncture the observer–system interface represents a possible pathway to solve the information loss paradox. The key point is to overcome the no-hair theorem for classical black holes. This states that the thermodynamics of black holes shall be described only in terms of three quantities: the mass, the spin and the electric charge of black holes, all the other classical degrees of freedom being irrelevant (Misner, Thorne & Wheeler 1973). The no-hair theorem can be avoided by resorting to the notion of quantum hairs. The latter are quantum numbers that black holes may carry, which are not associated with massless gauge fields and which may solve the information paradox, allowing for storing of information. This is a perspective that comes from an old idea (Coleman, Preskill & Wilczek 1992) that recently underwent a popular revival in the literature of high energy physics, thanks to the intuitions of Stephen Hawking (2015) and to the work of Hawking, Malcolm Perry and Andrew Strominger (2016) on soft photons. The very quantum-mechanical description of the theory of perception the authors move from can naturally encode quantum hairs. But then a first provocative question to address would be: Can an observer have access to quantum hairs and thus to the information that can be encoded in these latter entities? (Q1)

« 4 » Beyond this analogy between event horizon and interface of perception, it is possible to point out a more general correspondence between quantum degrees of freedom that are encoded on the observer-system boundary and some theories of quantum gravity that make explicit use of graph theory. Among the latter we mention loop quantum gravity (Rovelli 2004) and theoretical constructions that arise from string-nets (Levin & Wen 2005). Gauge interactions, and eventually also fundamental particles of the standard model (Bilson-Thompson, Markopoulou & Smolin 2007), can be derived in these two frameworks. The basic objects of these theories are graphs, namely sets of nodes interconnected by links, which are colored by fundamental representations of some continuous or discrete Lie group. These latter are sets of elements on which it is possible to define a product rule, recover a unit element and then find an inverse element that reproduces the unit element by virtue of the product rule. The redundancy that the authors propose to be deployed for unravelling the emergence of space, and in general spacetime, as an error-correcting code could be then associated with the irreducible representations that are assigned to the links of the graphs in these theories. This is exactly the same construction developed in loop quantum gravity or string-nets. The quantum states of the models of emergent spacetime are then recovered from the graphs that are taken into account. The colors, i.e., the irreducible representations of elements of the Lie group, are now associated with eigenvalues of the observable quantum operators of the theory. The dimension of the Hilbert space associated with the irreducible representations of a discrete or continuous group Lie group G – or eventually to a quantum group – that are assigned to the links of the graphs, is the natural instantiation of this line of thought and might represent redundancy.

« 5 » Moving from such an intuitive approach we are led to ask a second relevant question: How can we encode the observer-system duality in those models of quantum gravity that are phrased within the language of graphs (spin-networks, Wilson loops or string-nets)? (Q2) If we take into account the authors' analysis in (§§15–17), it seems natural to argue that the individuation of an interface distinguishes two subsystems of the Hilbert space, and thus implies that the total set of degrees of freedom encloses both the observer that perceives and the perceived system. Nonetheless, redundancy would require, in order to let emerge the notion of spacetime that satisfies the Einstein equivalence principle, that a continuous group structure G could be consistently defined (§§15–17) for the spaces of actions, and that this G could be connected to the Lorentz group.

« 6 » As the authors point out in §34, a sizable amount of energy expenditure is required for the holographic encoding, which is roughly proportional to the number of bits involved at the interface and to thermal energy for each degree of freedom. This implies that the redundancy increases in proportion to the dimension of the Hilbert space at the boundary between the observer and the system. Thus, the simple system described by a binary code, namely the Hilbert space of spin 1/2 particle in the physicist's jargon, might already turn into an extremely complicated model to be solved. Nonetheless, at least from a theoretical perspective, we may ask what happens if the Hilbert space at the boundary is composed by N degrees of freedom whose internal degeneracy is described by the irreducible representation of a Lie group G. The main last question I propose is therefore: What is the nature of these degrees of freedom at the interface between the observer and the system and what is the internal degeneracy group, namely the redundancy, connected to these degrees of freedom? (Q3)

« 7 » The answer to Q3 amounts to the correct reconstruction of the boundary physical theory. We must indeed recover the relevant degrees of freedom at the interface, and consistently describe at the quantum(-gravitational) level the interactions these undergo. The role of the symmetries, to which are connected charges that may play the role of bits, is indeed very intertwined with this aspect, as emphasized in a series of studies by Hawking, Perry and Strominger — see, e.g., Hawking, Perry & Strominger (2016).

« 8 » I wish also to emphasize that the role of quantum gravity is not only crucial to determine the dimensionality of the boundary Hilbert space – this pertains to the total set constituted by the "observer" and the "system," and accounts for the description of their interaction – but is relevant as well to regularize the maximum amount of degrees of freedom that shall be considered while reckoning the exchange of bits and the flow of information through the membrane. This provides a set-up in which we can operationally accomplish calculations, avoiding infinities.

« 9 » I end this brief commentary by recalling the authors' suggestive remark (in §38) – part of common belief in the community of quantum gravity that has been growing in recent years – that with the relation between interface's perception and holographic encoding we may only actually be probing the tip of an iceberg. A deeper understanding of the emergent nature of spacetime might indeed arise from the development of a theory of quantum information gravity that many authors are currently developing in the literature.

Antonino Marcianò joined as Associate Professor the Department of Physics at Fudan University in January 2014, becoming a member of the theory and high-energy division. Previously a post-doctoral researcher at Princeton University and Dartmouth College, he was studying models for cosmological inflation and CMBR physics, currently his main topics of research. In the USA, he also continued focusing on the Wilson-loop approach to Quantum Cosmology and Quantum Gravity, learnt while working at Aix-Marseille University, soon after his PhD at Sapienza University of Rome. His current research also encompasses the implementation in condensed-matter physics of mathematical tools borrowed from quantum gravity, as an attempt to address dynamics on lattice structures, including graphene, in non-perturbative regimes.

Received: 19 June 2017
Accepted: 26 June 2017

Authors' Response

Boundaries, Encodings and Paradox: What Models Can Tell Us About Experience

Chris Fields, Donald D. Hoffman, Chetan Prakash & Robert Prentner

> Upshot • Formal models lead beyond ordinary experience to abstractions such as black holes and quantum entanglement. Applying such models to experience itself makes it seem unfamiliar and even paradoxical. We suggest, however, that doing so also leads to insights. It shows, in particular, that the "view from nowhere" employed by the theorist is both essential and deeply paradoxical, and it suggests that experience has an unrecorded, non-reportable component in addition to its remembered, reportable component.

« 1 » We thank our commentators for their insightful criticism. While each of them chooses a different focus for their comments, the issues they raise overlap considerably. We highlight in what follows what we take to be the major issues, and attempt to show how they relate both to what we propose in the target article and to one another.

The "classical world" is the explanandum

« 2 » Constructivists, phenomenologists, and others who reject naive realism are faced with the task of explaining a sharable experience of a classical world – a world of "tables and chairs and dogs and cats and people" (**Eric Dietrich** §1). Even the "naturalized" sciences, however, face this challenge. This is obvious in the case of quantum theory, but even the classical theory of atom-based matter – the classical physics of the late 19th century – faces the problem of how clouds of atoms could appear to us to be tables or chairs. It is less obvious in the case of biology and psychology, but here it must be explained how agglomerations of cells – i.e., organisms – could self-assemble in ways that allow the experience of such things as tables and chairs as opposed to, say, just brightness and saltiness.

« 3 » We agree with **Dietrich** (§14) that the *experience* of a classical world is ineluctable. When we open our eyes, we see bounded objects with definite shapes, sizes and locations; when we open our ears we hear tones with definite loudness and pitch. Our goal is to explain why we have such experiences. **Dietrich** suggests that the experience of a classical world is ineluctable because there is an ontologically real classical world, one with a "mind-independent spacetime" that is "filled with mind-independent objects" (§2). We "visit" this world by opening our eyes and ears. According to **Dietrich** (§12), an utterly differently structured quantum world that we can access (since the 1920s) only via our thoughts can be considered to be equally real, and there may be other equally real worlds with yet different structures that we cannot access at this time. From a constructivist perspective, these "worlds" are all constructs, one of our perceptual systems and the other(s) of our theoretical imaginations. Why the former should provide compelling experimental evidence for the latter remains a mystery. Why we can only express our theories – even to ourselves, in thoughts – using classical symbols is also mysterious.

« 4 » We attempt to address these questions by appealing to a specific mechanism: holographic encoding on an interface that employs spacetime as an error-correcting code. We (each) see a classical world, in our view, because we (each) have this kind of interface. The "objects" – including objects of thought – that our interfaces present to us are eigenforms. As Heinz von Foerster (1976) emphasized, eigenforms and the corresponding eigenbehaviors are (at least approximate) fixed points of multiply repeated (ideally infinitely repeated) perception-action loops (cf. **Louis Kauffman**'s commentary). Eigenform and eigenbehavior must be classically correlated across these repetitions; hence the process of repetition, whether it is conscious or not, constitutes a *memory*. It is this memory of classical correlation that confers classicality on the "classical world" of our interface-encoded experience.

« 5 » If we are correct, the "classical world" is not a world at all, but is *only* an experience. The classical-world experience is ineluctable because the interface that encodes it is the only interface we have; as Ernst von Glasersfeld puts it, summarizing three millennia of philosophical empiricism, "it is impossible to compare our image of reality with a reality outside" (Glasersfeld 1981: 89). When we imaginatively construct theories of what lies beyond the interface, we construct and express them using symbols and diagrams that our interfaces allow: classical symbols and diagrams that have definite arrangements and shapes. Such symbols and diagrams are, like our percepts, eigenforms, fixed points that are only recognizable through repeated use. We have no choice in our use of *classical* symbols and diagrams, as our experiences of theory construction and our experiences of our constructed theories are *experiences* and so are encoded on our interfaces. The classical symbols and diagrams that we use to express our theories make use of redundancy in space and time; hence they enable error correction.

« 6 » What we have called the classical world*view*, on the other hand, is an assumption that the classical world of our experience is not just encoded on our interfaces, but also exists beyond them as an ontologically real structure comprising a multitude of well-defined, bounded, time-persistent macroscopic objects. We see tables and chairs, in this worldview, because tables and chairs (not just clouds of atoms) are out there, bouncing light into our eyes. Perception is (mostly) veridical because the interfaces through which we have perceptual experiences are (mostly) transparent. The world, on this worldview, is not a black box at all, but rather a (mostly) white one. What you see is what you get. **Dietrich** argues (§3, §12) that this world/worldview distinction is illegitimate without empirical evidence that our model is correct. We disagree: the classical worldview is an explicit philosophical claim or, more commonly, an implicit and perhaps innate assumption that can be (and in point of fact is) made independently of whether the classical world that it postulates actually has the ontological status that the classical worldview claims it to have. On the other hand, we agree with **Dietrich** that there is a deep issue here: stating this distinction is making a *statement*, and making any particular statement is a classical act. If the classical worldview is rejected, the status of statements is cast into doubt; it is unclear how anyone could speak one particular sentence or think one particular thought. Memory and communication both become paradoxical. Any non-classical theory seems to require, as Niels Bohr argued, a classical metatheory just to support *language*. Here a dialetheic world (**Dietrich** §12) seems inescapable (Dietrich & Fields 2015).

« 7 » While we do not, as **Dietrich** points out, have direct empirical evidence for our model, there is plentiful (albeit indirect) evidence for holography as a mechanism (see, e.g., **Antonio Marcianò**'s commentary). Many would argue, moreover, that the mounting evidence for quantum effects at macroscopic scales demonstrates empirically that the classical worldview is wrong. As **Dietrich** emphasizes, accepting this argument requires the acceptance of another deep paradox. Experiments, in particular, require time-persistent observers and apparatuses that interact while remaining separable in the physicist's sense of having independently characterizable states. Joint states of interacting systems are not, however, separable under the unitary evolution prescribed by quantum theory. This paradox can be stated starkly: local decoherence requires global coherence, i.e., global entanglement. From a global quantum-theoretic perspective, both decoherence and the classical world it produces are epiphenomenal.

« 8 » **Dietrich** also points out (§14) that we have offered no theory of how human beings can formulate, within their classical interfaces, theories of the non-classical. This is a fair challenge that we hope someday to accept.

Consciousness is fundamental, but architecture must be fundamental too

« 9 » Both **Dietrich** (§13) and **Urban Kordeš** (§10) suggest that we are trying to explain phenomenal consciousness, or are at any rate not taking it to be fundamental. We were perhaps not sufficiently clear that we take phenomenal consciousness to be fundamental and irreducible, and simply assume that conscious agents have it. However, we also assume that conscious agents have an architecture in addition to consciousness. The structure and content of phenomenal consciousness (i.e., experience) alone is, we claim, insufficient to explain itself, e.g., insufficient to explain the structure and content of the experienced classical world.

« 10 » **Kordeš** specifically argues that our distinction between the experience space *X* and the space *G* of available actions is a mistake; *G*, **Kordeš** suggests, should be a subset, presumably a proper one, of *X*. "Being autonomous," he claims, "means that the agent chooses from the options the agent itself constructs rather than from pre-given options" (§11). Placing *G* within *X* results, however, in an agent aware of every available action and of every choice of action. No actions by such an agent can be "automatic" as psychologists such as John Bargh and Tanya Chartrand (1999) use this term. Genuine autonomy, moreover, requires that the agent be able to actually perform whatever action is chosen. This is possible only if the world never interferes to prevent a chosen action. The conscious agent (CA) formalism separates *G* from *X* not just to enable automaticity, but also to take the evident ability of the world to interfere with our desires into account. The best argument for the existence of a world independent of your own mind is, as The Rolling Stones explain it, "you can't always get what you want."

« 11 » Postulating an architecture is, by its very nature, going beyond "lived experience" to the realm of theoretical models. We fully agree with **Kordeš** that pretending to "eyes of God" that "[see] all agents, their actions and interactions" (§13) is a mistake, but we nonetheless regard an ability to build, consider, and derive predictions from theoretical models as an essential adjunct to phenomenology. The formalism and diagrams of von Foerster, for example, compose such a model, as do those of Karl Friston or Wojciech Zurek or indeed of any other author who claims to explain or predict any experience of any observer. **Kordeš** is no exception. "By renouncing the view from nowhere, consciousness appears everywhere" (**Kordeš** §15) may well be a report of first-person experience, but saying *how this happens* requires a model. For many, moreover, consciousness appears everywhere only *from* a theoretical, view-from-nowhere perspective, one from which the futility of attempts to make consciousness "emerge" from something else becomes evident.

« 12 » Consciousness appears everywhere in the CA framework via a postulate: conscious realism (see §9 in our target article). This postulate is not as radical as it seems. Two CAs defined to have the same "world" set *W* can be taken to represent two "points of view" on *W*. If, however, *W* is reconceptualized as simply the *information channel* via which the agents interact, its degrees of freedom can be subsumed into the perception and action maps of the agents to produce the interacting-agent configuration shown in Figure 1c of our target article. From the perspective of either agent, the "world" is indistinguishable from the other agent. René Descartes realized this in his *Meditations*, stating that nothing in his experience could prove that he was not interacting with an "evil demon" that synthesized his every percept. The currently fashionable idea that we (each) live in a computer simulation constructed by some advanced race, maybe even our own descendants (Bostrom 2003), updates Descartes. The simulation is, in this view, the channel by which the aliens, or maybe our grand^n-children, toy with us.

The interface is a boundary in state space, not spacetime

« 13 » **Kauffman** and **Konrad Werner** both wonder how the interface is defined, a question that is present but implicit for both **Dietrich** and **Kordeš**. Kauffman asks, in particular, (Q2) whether we require the interface to be a "physical surface," later attributing to us the notion that "the fundamental source of the epistemic boundary is spacetime itself" (§7). The word "physical" here is ambiguous; physicists often use it to mean merely "consistently describable in the language of physics," ruling out as "unphysical" only situations with mathematical descriptions that are self-contradictory or meaningless. We can, however, state categorically that we do not require the interface to be a boundary in spacetime, and we apologize if anything in our text suggests this. We regard spacetime as a way of encoding information on an interface, one that may or may not be used, but that provides the benefit of some level of error correction. Human experience and thus the (typical) human interface employs spacetime to advantage for encoding percepts, some concepts (e.g., those of geometry), and much of what we imagine, but other kinds of observers may have interfaces that do not employ spacetime, or that employ spacetimes with more or fewer dimensions or even different geometries from ours. Encodings of some kinds of human experience, e.g., of emotions or epistemic feelings, tend to employ time but not space. Nothing requires or even suggests a common encoding across the entire interface.

« 14 » The notions of open and closed boundaries of classical mereotopology are motivated by the characteristics of ordinary objects occupying continuous, locally Euclidean spacetime. Hence it is unsurprising that, as **Werner** shows, they are of little use in understanding the kind of interface proposed here. **Werner** rejects, in particular, our characterization of observer (or "perceiver") and environment ("outside world") as mereotopologically neither open nor closed (**Werner** §5). If either is open, its complement must be closed (Smith 1996). Observer and environment are, however, on this model entirely equivalent and interchangeable; this is why we draw them symmetrically and prefer the neutral "Alice" and "Bob" nomenclature to the connotation-laden "observer" and "environment." Nothing motivates any structural distinction between the two; hence there is no justification for a mereotopological distinction. Given that they interact, we are left with the situation that **Werner** (§4) labels "1B*pw*": both systems are closed and they share a boundary. While the boundary is shared, however, the systems cannot both be closed: observer and environment together compose the entire universe, which, as Barry Smith (1996) points out, is boundaryless and hence not mereotopologically closed (it is, however, closed in the physicist's sense of not interacting with anything). This situation is rendered even more paradoxical by noting that observer and environment each appears fully embedded in the other when viewed from their own perspective.

« 15 » **Kauffman** remarks that "the most generally applicable epistemic boundary is any distinction whatsoever" (§6). The distinctions between red and green or between happy and sad are examples. Any property that supports such a distinction (what physicists call a "degree of freedom") can be thought of as a component of the *state* of a system. The boundaries in which we are interested are boundaries in the abstract state space (as **Kauffman** §7 points out, this is a

Hilbert space in quantum theory) of the universe. Observer and environment are distinguished as subsystems by the states that they can occupy. The epistemic boundary between them – the boundary by which *we*, as theorists, distinguish them – is their shared interface. The states on this boundary are available to encode experiences; they implement the respective spaces X of observer and environment in the CA formalism. *What is encoded* on the interface at any instant of either system-relative time depends on how the two systems are interacting at that time. The interaction need not involve spatial degrees of freedom, as **Kauffman** makes clear in his discussion of entanglement (§7).

All boundaries encode experience, but all boundaries can be erased

« 16 » **Kauffman**'s remark that "any distinction whatsoever" creates an epistemic boundary is, however, even more powerful than this. It implies, when taken seriously, that every possible boundary in state space encodes experience. Every system is an observer; likewise, every system is an observed environment. Every state corresponds to an experience on some interface. The universe is, therefore, filled with experiencers and filled with experience. In this sense, contra **Kordeš** (§17), the abstract space in which agents live is indeed a space of phenomenal experience. Each agent, however, experiences only what is encoded on its own interface. Sensations, thoughts, feelings, imaginations, the experiences of deciding or doing, all are encoded on the interface. All are eigenforms. Each agent's internal, "bulk" states are experientially inaccessible to it, even though each of them is on the interface of and hence encodes accessible experience for *some* agent. To see this in the simpler arena of spacetime, think of the constant experiences of your own neurons (of which Cook 2008 provides a compelling description), all of which are inaccessible to you.

« 17 » Expanding one's (theoretical) perspective to the entire universe considered as a whole, however, produces not **Kauffman**'s hoped-for abduction but **Dietrich**'s dialetheic paradox. As described in §7 of our target article, both classical and quantum physics allow inter-system boundaries to be moved or erased arbitrarily without affecting joint-system dynamics (e.g., Zanardi 2001; Dugíc & Jekníc-Dugíc 2008; Harshman & Ranade 2011); this constancy of whole-system dynamics under arbitrary decomposition has been termed "decompositional equivalence" (Fields 2016). Within the CA formalism, decompositional equivalence is implemented by the arbitrary composability of Markov processes. The universe as a whole has no "outer" boundary; decompositional equivalence allows the erasing of any "inner" boundaries as well. Hence the universe can be considered to be filled with observers and experiences as described above, but the boundaries defining these observers can also be erased with no effect. In the CA formalism, the universe can be considered to be a CA or any combination of CAs, but it can also be considered to be a single set W mapped to itself. If any distinction creates a boundary, such a boundaryless system can make no distinctions. With no boundary to serve as an interface and no ability to make distinctions, the universe has no experience space X and no experiences. It has no point of view, on itself or on anything else. John Wheeler's well-known statement (**Kauffman** §3) is, therefore, misleading. The universe is composed of observer-participants, but is, when viewed as a boundaryless whole, itself neither an observer nor a participant.

« 18 » Taking actions into account deepens the above paradox. Boundaries encode not just experiences but actions: the perceptions of each agent are the actions of its environment and vice versa. The actions of agents drive the evolution of the universe; the dynamics of a universe entirely composed of agents is nothing beyond the combination of all of their actions. Yet from the (theoretical) perspective of the entire universe, none of the boundaries matters. Decompositional equivalence allows the erasing of all boundaries with no effect. From the perspective of the whole universe, there is no spacetime (indeed no classical information) and nothing is happening. The universe is in a pure entangled state. That this fixed point exists is the physical content of the Wheeler-DeWitt equation.

« 19 » The paradox posed by the "universal view" is, however, deeper still. The boundary erasure allowed by decompositional equivalence erases all interfaces and hence all encoded experience. From the (theoretical) perspective of the entire universe, consciousness and its contents are, like decoherence, epiphenomenal. Decompositional equivalence renders a universe filled with awareness and a universe containing no awareness indistinguishable from a (theoretical) perspective that stands "outside" of it. The "view from nowhere," even when adopted via an abstract model, is inherently paradoxical.

Experience is both classical and non-classical

« 20 » A partial resolution of this paradox of disappearing awareness may come from an unlikely corner. **Marcianò** focuses on a particular system for which the state-space boundary corresponds to a spatial boundary, the black hole, and asks (Q1) how our approach might deal with the paradox that black holes appear to destroy information whenever they gain energy, in violation of quantum theory's requirement of unitarity and hence information conservation. As **Marcianò** points out, one answer to this paradox is to recognize that black holes are only apparently classical objects; they are entangled with the rest of the universe by "soft" photons and possibly other "quantum hair" (see Strominger 2017 for a recent elaboration of this view).

« 21 » As all systems smaller than the universe as a whole are observers in our approach, black holes are observers. Indeed, they are *ideal* observers: all information (particles or waves) that contacts their surfaces is both fully absorbed and holographically encoded. Black holes are also ideal actors: they constantly alter the states of their environments by emitting Hawking radiation. These observations and actions are classical: they can be observed by (i.e., can encode information on the interface of) an external observer. When the situation is viewed quantum-mechanically, however, on the two sides of a black hole's boundary are simply quantum states, which to preserve unitarity must be entangled. The correlations that implement this entanglement cross the boundary; they are the soft quantum hairs. In Andrew Strominger's formalism, these soft hairs are the decohering environment for the Hawking radiation; the latter is detectable by us only because the soft hairs are there. The soft hairs themselves, however, are not detectable; they carry zero energy and hence cannot encode classical bits on an interface.

« 22 » The interfaces of black holes, our ideal observers, are thus more complicated than is depicted in Figure 3 of our target article. Not only do they encode classical information; they are also a locus of quantum correlation. The former cannot happen without the latter. If the encoded classical information is the content of *recallable, reportable, classical* experience, the kind of experience that can be remembered or put into a sentence, then it is natural to regard the boundary-crossing non-classical correlations as a kind of ineffable, non-classical experience that can be neither remembered nor reported. Without this ineffable experience, recallable, reportable experience could not occur.

« 23 » If all of the boundaries in the universe are erased, the classical, reportable experience disappears. It is, as noted earlier, epiphenomenal from a whole-universe perspective. The *non-classical experience*, however, remains. The quantum correlations that implement this non-classical experience constitute the universal entangled state, the fixed point of the universe's timeless evolution. Hence **Kauffman**'s abduction can be partially recovered: the universe remains filled with ineffable, non-classical experience even when all observer–system boundaries have been removed. Perhaps **Kauffman**'s "places of ambiguity" (§4) point to this non-classical experience as surely as do **Dietrich**'s dialetheia. William James's (1892) "fringe" of consciousness similarly seems to point here.

"What is it like?" is not one question but two

« 24 » **Kordeš** (§14) introduces the traditional distinction between what consciousness *does* and what it *is like*, suggesting that we may address the former but can say nothing about the latter. We disagree, for we claim that "what is it like?" is two distinct questions. One asks what *sorts of experiences* might we expect a system to have, while the other asks what each of those experiences is like for each system that has it. The first of these questions can be answered, maybe not in all cases, but in some. We can expect bacteria, for example, to experience saltiness and expect humans to experience time-persistent objects located in 3D space. We can expect both to experience the difference between well-being and its absence (Peil Kauffman 2015). What these experiences are like for each individual experiencer, however, remains unanswerable. It remains unanswerable, we would argue, even from a first-person perspective. What is the experience of green like? It is like green! Even elaborating, saying that green is more like cyan than red, contributes nothing to capturing in non-experiential terms the experience of greenness. Remembering and then describing the greenness makes it, if anything, less immediate and vivid. Forcing experience into language, even first-person language, distances it.

« 25 » Holography provides a mechanism for rendering experience classical. Beyond that, answering the "what sorts of experiences?" question requires the investigation and modeling of the particular interfaces of particular kinds of systems – e.g., particular kinds of organisms – or even particular individuals. It requires us to take **Werner**'s questions (§14) about the structures of sensory and cognitive systems seriously. Such questions inevitably lead to the field station, the laboratory, or the clinic. It is, once again, a fair challenge to ask how and even if such investigations can be fully and adequately described within a purely constructivist framework. We doubt it.

« 26 » Framed in **Marcianò**'s terms, "what sorts of experiences?" becomes a mathematical question about the formal structures of model interfaces. Given an observer-environment pair, for example, what group structures characterize their interface (Q3)? We have addressed this question from the reverse direction, showing that an interface with a given group structure imposes that structure on the experienced world (Hoffman, Singh & Prakash 2015 and current work). For a finite interface and hence a finite classical experience space X, such groups are finite; hence they can at best approximate continuous group transformations, e.g., those of the Lorentz group (Q2). Whether the CA formalism can replicate the graph structures employed by physicists while maintaining its intended interpretation is a topic of ongoing investigation.

« 27 » **Kauffman** raises a general question about encoded experiences: what does it mean to say that the informational redundancy enabled by spacetime or any other group structure corrects errors (Q1)? As **Kauffman** notes, an experience *per se* simply is what it is; there is no sense in calling it an "error." The errors that are corrected, in our view, are errors of association between experiences and actions. Depth perception, for example, enables accurate grasping; disrupting depth perception introduces errors. In some cases, experience-action associations are mediated by intervening experiences. An accurate representation of the time between a current sensory experience and a remembered experience – as encoded in an experience of recall happening now – may be required to choose an appropriate action, e.g., whether to hurry to avoid being late. It is errors of this sort that can decrease fitness, and in extreme cases send fitness toward zero, stopping further input. For an organism, no action is repeated *ad infinitum* and no eigenform is stable forever. In a universe where you cannot always get what you want, you are better off having an interface that gets you what you need.

Received: 4 July 2017
Accepted: 7 July 2017

Combined references

Adams R. A., Friston K. J. & Bastos A. M. (2015) Active inference, predictive coding and cortical architecture. In: Casanova M. F. & Opris I. (eds.) Recent advances in the modular organization of the cortex. Springer, Berlin: 97–121.

Arkani-Hamed N. & Trnka J. (2014) The amplituhedron. Journal of High Energy Physics 2014: 30.

Ashby W. R. (1956) Introduction to cybernetics. Chapman and Hall, London.

Baillargeon R. (2008) Innate ideas revisited: For a principle of persistence in infants' physical reasoning. Perspectives on Psychological Science 3(1): 2–13.

Bargh J. A. & Chartrand T. L. (1999) The unbearable automaticity of being. American Psychologist 54(7): 462–479.

Bateson G. (1987) Steps to an ecology of mind. Jason Aronson, London. Originally published in 1972.

Bekenstein J. D. (1973) Black holes and entropy. Physical Review D 7(8): 2333–2346.

Bennett C. H. (2003) Notes on Landauer's Principle, reversible computation, and Maxwell's Demon. Studies in the History and Philosophy of Modern Physics 34(3): 501–510.

Bilson-Thompson S. O., Markopoulou F. & Smolin L. (2007) Quantum gravity and the standard model. Classical and Quantum Gravity 24(16): 3975–3993.

Blackmore S. (2013) Consciousness: An introduction. Routledge, London.

Bohr N. (1928) The quantum postulate and the recent developments of atomic theory. Nature 121: 580–590.

Bohr N. (1958) Atomic physics and human knowledge. Wiley, New York.

Bostrom N. (2003) Are we living in a computer simulation? Philosophical Quarterly 53(11): 243–255.

Bousso R. (2002) The holographic principle. Reviews of Modern Physics 74(3): 825–874.

Chalmers D. (1995) Facing up to the problem of consciousness. Journal of Consciousness Studies 2(3): 200–219.

Clark A. & Chalmers D. (1998) The extended mind. Analysis 58: 10–23.

Coleman S. R., Preskill J. & Wilczek F. (1992) Quantum hair on black holes. Nuclear Physics B 378(1–2): 175–246.

Conway J. & Kochen S. (2006) The free will theorem. Foundations of Physics 36(10): 1441–1473.

Cook N. D. (2008) The neuron-level phenomena underlying cognition and consciousness: Synaptic activity and the action potential. Neuroscience 153(3): 556–570.

Csibra G. & Gergely G. (2012) Teleological understanding of actions. In: Banaji M. R. & Gelman S. A. (eds.) Navigating the social world: What infants, children and other species can teach us. Oxford University Press, Oxford: 38–43.

Dietrich E. & Fields C. (2015) Science generates limit paradoxes. Axiomathes 25(4): 409–432.

Dietrich E. & Hardcastle V. (2005) Sisyphus's boulder: Consciousness and the limits of the knowable. John Benjamins, New York.

Dugíc M. & Jeknı́c-Dugíc J. (2008) What is "system": The information-theoretic arguments. International Journal of Theoretical Physics 47(3): 805–813.

D'Ariano G. M. & Perinotti P. (2017) Quantum cellular automata and free quantum field theory. Frontiers in Physics 12: 120301.

Eibenberger S., Gerlich S., Arndt M., Mayor M. & Txen J. (2013) Matter-wave interference of particles selected from a molecular library with masses exceeding 10,000 amu. Physical Chemistry and Chemical Physics 15: 14696–14700.

Fields C. (2012) A model-theoretic interpretation of environment-induced superselection. International Journal of General Systems 41(8): 847–859.

Fields C. (2013) A whole box of Pandoras: Systems, boundaries and free will in quantum theory. Journal of Experimental and Theoretical Artificial Intelligence 25(3): 291–302.

Fields C. (2014) Consistent quantum mechanics admits no mereotopology. Axiomathes 24(1): 9–18.

Fields C. (2016a) Building the observer into the system: Toward a realistic description of human interaction with the world. Systems 4: 32.

Fields C. (2016b) Decompositional equivalence: A fundamental symmetry underlying quantum theory. Axiomathes 26(3): 279–311.

Foerster H. von (1960) On self-organizing systems and their environments. In: Yovits M. C. & Cameron S. (eds.) Self-organizing systems. Pergamon, London: 31–50. Reprinted in: Foerster (2003: 1–19)
▸ http://cepa.info/1593

Foerster H. von (1970) Thoughts and notes on cognition. In: Gavin P. (ed.) Cognition: A multiple view. Spartan, New York: 25–48. Reprinted in: Foerster (2003: 169–189)
▸ http://cepa.info/1637

Foerster H. von (1973) On constructing a reality. In: Preiser F. E. (ed.) Environmental design research. Volume II. Dowden, Hutchinson & Ross, Stroudsburg PA: 35–46. Reprinted in: Foerster (2003: 211–227)
▸ http://cepa.info/1278

Foerster H. von (1976) Objects: Tokens for (eigen-) behaviors. ASC Cybernetics Forum 8(3–4): 91–96. Reprinted in Foerster H. von (2003) Understanding understanding. Springer, New York: 261–271.
▸ http://cepa.info/1270

Foerster H. von (1979) Cybernetics of cybernetics. In: Krippendorf K. (ed.) Communication and control. Gordon & Breach, New York: 5–8. Reprinted in: Foerster (2003: 283–286)
▸ http://cepa.info/1707

Foerster H. von (1981) Notes on an epistemology for living things. In: Foerster H. von, Observing Systems. Intersystems Publications, Seaside CA: 258–271.

Foerster H. von (1984) On constructing a reality. In: Watzlawick P. (ed.) The invented reality. W. W. Norton, New York: 41–62.

Foerster H. von (2003) Understanding understanding. Springer, New York.

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

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

Friston K., Levin M., Sengupta B. & Pezzulo G. (2015) Knowing one's place: A free-energy approach to pattern regulation. Journal of the Royal Society: Interface 12: 20141383.

Fuchs C. (2010) QBism: The perimeter of quantum Bayesianism. arXiv: 1003.5209.

Fuchs C. A. & Stacey B. C. (2016) Some negative remarks on operational approaches to quantum theory. In: Chiribella G. & Spekkens R. W. (eds.) Quantum theory: Informational foundations and foils. Springer, Berlin: 283–305.

Geisler W. S. & Diehl R. L. (2003) A Bayesian approach to the evolution of perceptual and cognitive systems. Cognitive Science 27(3): 379–402.

Ghirardi G. C., Rimini A. & Weber T. (1986) Unified dynamics for microscopic and macroscopic systems. Physical Review D 34(2): 470–491.

Gibson J. J. (1979) The ecological approach to visual perception. Houghton-Mifflin, Boston.

Glanville R. (1982) Inside every white box there are two black boxes trying to get out. Behavioral Science 27(1): 1–11. ▸ http://cepa.info/2365

Glasersfeld E. von (1981) The concepts of adaptation and viability in a constructivist theory of knowledge. In: Sigel I. E., Brodzinsky D. M. & Golinkoff R. M. (eds.) New directions in Piagetian theory and practice. Erlbaum, Hillsdale NJ: 87–95. ▸ http://cepa.info/1357

Goodale M. A. & Milner A. D. (1992) Separate visual pathways for perception and action. Trends in Neurosciences 15(1): 20–25.

Harshman N. L. & Ranade K. S. (2011) Observables can be tailored to change the entanglement of any pure state. Physical Review A 84(1): 012303.

Hawking S. W. (2015) The information paradox for black holes. arXiv: 1509.01147.

Hawking S. W., Perry M. J. & Strominger A. (2016) Soft hair on black holes. Physical Review Letters 116(23): 231301.

Hensen B., Bernien H., Dréau A. E., Reiserer A., Kalb N., Blok M. S., Ruitenberg J., Vermeulen R. F. L., Schouten R. N., Abellán C. & Amaya W. (2015) Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. Nature 526: 682–686.

Hickok G. & Poeppel D. (2007) The cortical organization of speech processing. Nature Reviews Neuroscience 8: 393–402.

Hoffman D. D. & Prakash C. (2014) Objects of consciousness. Frontiers in Psychology 5: 577.

Hoffman D. D., Singh M. & Prakash C. (2015) The interface theory of perception. Psychonomic Bulletin & Review 22(6): 1480–1506.

Husserl E. (1982) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. First book: General introduction to a pure phenomenology. Translated by Fred Kersten. Nijhoff, The Hague. German original published in 1913.

Husserl E. (2012) Ideas: General introduction to pure phenomenology. Routledge, London. German original published in 1913.

Initials? → Misner C., Thorne K. & Wheeler J. (1973) Gravitation. W. H. Freeman, San Francisco CA.

James W. (1892) Psychology. University of Notre Dame Press, Notre Dame IN. Reprinted 2001 by Dover Press, Mineola NY.

Kauffman L. H. (2003) Eigenforms – Objects as tokens for eigenbehaviors. Cybernetics and Human Knowing 10(3–4): 73–90. ▸ http://cepa.info/1817

Kauffman L. H. (2009) Reflexivity and eigenform: The shape of process. Constructivist Foundations 4(3): 121–137. ▸ http://constructivist.info/4/3/121

Kauffman L. H. (2011) Eigenforms and quantum physics. Cybernetics and Human Knowing 18(3–4): 111–121. ▸ http://cepa.info/1818

Kordeš U. (2016a) Going beyond theory. Constructivist Foundations 11(2): 375–385. ▸ http://constructivist.info/11/2/375

Kordeš U. (2016b) Where is consciousness? Constructivist Foundations 11(3): 552–554.

Landauer R. (1961) Irreversibility and heat generation in the computing process. IBM Journal of Research and Development 5(3): 183–195.

Landauer R. (1999) Information is a physical entity. Physica A 263(1–4): 63–67.

Landsman N. P. (2007) Between classical and quantum. In: Butterfield J. & Earman J. (eds.) Handbook of the philosophy of science: Philosophy of physics. Elsevier, Amsterdam: 417–553.

Levin M. A. & Wen X.-G. (2005) String-net condensation: A physical mechanism for topological phases. Physical Review B 71(4): 045110.

Libet B., Gleason C. A., Wright E. W. & Pearl D. K. (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness–potential) Brain 106(3): 623–642.

Luo Y. & Baillargeon R. (2010) Toward a mentalistic account of early psychological reasoning. Current Directions in Psychological Science 19(5): 301–307.

Mandelbrot B. (1982) The fractal geometry of nature. Freeman, San Francisco.

Manning A. G., Khakimov R. I., Dall R. G. & Truscott A. G. (2015) Wheelers' delayed-choice gedanken experiment with a single atom. Nature Physics 11: 539–542.

Mark J. T., Marion B. B. & Hoffman D. D. (2010) Natural selection and veridical perceptions. Journal of Theoretical Biology 266(4): 504–515.

Marr D. (1982) Vision. Freeman, San Francisco.

Martin A. (2007) The representation of object concepts in the brain. Annual Review of Psychology 58: 25–45.

Maturana H. R. & Varela F. J. (1980) Autopoiesis and cognition: The realization of the living. Reidel, Boston.

Moore E. F. (1956) Gedankenexperiments on sequential machines. In: Shannon C. W. & McCarthy J. (eds.) Automata studies. Princeton University Press, Princeton NJ: 129–155.

Nagel T. (1989) The view from nowhere. Oxford University Press, Oxford.

Neumann J. von (1955) Mathematical foundations of quantum mechanics. Princeton University Press: Princeton NJ. German originally published in 1935.

Ollivier H., Poulin D. & Zurek W. H. (2004) Objective properties from subjective quantum states: Environment as a witness. Physical Review Letters 93: 220401.

Ollivier H., Poulin D. & Zurek W. H. (2005) Environment as a witness: Selective proliferation of information and emergence of objectivity in a quantum universe. Physical Review A 72: 042113.

Palmer S. E. (1999) Vision science: Photons to phenomenology. MIT Press, Cambridge MA.

Pastawski F., Yoshida B., Harlow D. & Preskill J. (2015) Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence. Journal of High Energy Physics 2015: 149.

Pattee H. H. (2001) The physics of symbols: Bridging the epistemic cut. Biosystems 60(1–3): 5–21.

Pearl J. (1988) Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann, San Mateo CA.

Peil Kauffman K. (2015) Emotional sentience and the nature of phenomenal experience. Progress in Biophysics and Molecular Biology 119(3): 545–562.

Penrose R. (1996) On gravity's role in quantum state reduction. General Relativity and Gravitation 28(5): 581–600.

Pizlo Z., Li Y., Sawada T. & Steinman R. M. (2014) Making a machine that sees like us. Oxford University Press, New York.

Quine W. V. O. (1960) Word and object. MIT Press, Cambridge MA.

Quine W. V. O. (1970) On the reasons for the indeterminacy of translation. Journal of Philosophy 67(6): 178–183.

Riegler A. (2012) Constructivism. In: L'Abate L. (ed.) Paradigms in theory construction. Springer, New York: 235–255. ▸ http://cepa.info/446

Roederer J. (2005) Information and its role in nature. Springer, Berlin.

Rovelli C. (2004) Quantum gravity. Cambridge University Press, Cambridge.

Rubino G., Rozema L. A., Feix A., Araújo M., Zeuner J. M., Procopio L. M., Brukner Č. & Walther P. (2017) Experimental verification of an indefinite causal order. Science Advances 3(3): E1602589.

Saini A. & Stojkovic D. (2015) Radiation from a collapsing object is manifestly unitary. Physical Review Letters 114: 111301.

Schlosshauer M. (2006) Experimental motivation and empirical consistency of minimal no-collapse quantum mechanics. Annals of Physics 321(1): 112–149.

Schlosshauer M. (2007) Decoherence and the quantum to classical transition. Springer, Berlin.

Smith B. (1996) Mereotopology: A theory of parts and boundaries. Data and Knowledge Engineering 20(3): 287–303.

Smith J. E. & Nair R. (2005) The architecture of virtual machines. IEEE Computer 38(5): 32–38.

Spelke E. S. (1994) Initial knowledge: Six suggestions. Cognition 50(1–3): 431–445.

Spencer Brown G. (1969) Laws of form. Allen & Unwin, London.

Strominger A. (2017) Black hole information revisited. Preprint arxiv: 1706.07143v1 (hep-th)

Susskind L. (1995) The world as a hologram. Journal of Mathematical Physics 36: 6377–6396.

Susskind L. (2016) Computational complexity and black hole horizons. Fortschritte Physik 64(1): 24–43.

Swingle B. (2012) Entanglement renormalization and holography. Physical Review D 86: 065007.

Tarski A. (1944) The semantic conception of truth and the foundations of semantics. Philosophy and Phenomenological Research 4(3): 341–376.

Tegmark M. (2012) How unitary cosmology generalizes thermodynamics and solves the inflationary entropy problem. Physical Review D 85: 123517.

Thompson D'A. W. (1945) On growth and form. Cambridge University Press, Cambridge. Originally published in 1917.

't Hooft G. (1993) Dimensional reduction in quantum gravity. In: Ali A., Ellis J. & Randjbar-Daemi S. (eds.) Salamfestschrift. World Scientific, Singapore: 284–296.

Trivers R. L. (2011) The folly of fools. Basic Books, New York.

Varela F. J., Thompson E. & Rosch E. (1991) The embodied mind: Cognitive science and human experience. MIT Press, Cambridge MA.

Wallace D. (2008) Philosophy of quantum mechanics. In: Rickles D. (ed.) The Ashgate companion to contemporary philosophy of physics. Ashgate, Aldershot: 16–98.

Wegner D. M. (2003) The mind's best trick: How we experience conscious will. Trends in Cognitive Sciences 7(2): 65–69.

Weinberg S. (2012) Collapse of the state vector. Physical Review A 85: 062116.

Wheeler J. A. (1990) Information, physics, quantum: The search for links. In: Zurek W. (ed.) Complexity, entropy and the physics of information. Westview, Boulder CO: 3–28.

Wigner E. P. (1962) Remarks on the mind-body question. In: Good I. J. (ed.) The Scientist Speculates. Basic Books, New York: 284–302.

Winograd T. & Flores F. (1986) Understanding computers and cognition: A new foundation for design. Ablex Publishing, Norwood.

Zahavi D. (2004) Phenomenology and the project of naturalization. Phenomenology and the cognitive sciences 3(4): 331–347. ▸ http://cepa.info/2375

Zanardi P. (2001) Virtual quantum subsystems. Physical Review Letters 87(7): 077901.

Zeh D. (1970) On the interpretation of measurement in quantum theory. Foundations of Physics 1(1): 69–76.

Zeh D. (1973) Toward a quantum theory of observation. Foundations of Physics 3(1): 109–116.

Zurek W. H. (2003) Decoherence, einselection, and the quantum origins of the classical. Reviews of Modern Physics 75(3): 715–775.