Eigenforms, interfaces and holographic encoding

Toward an evolutionary account of objects and spacetime

Chris Fields • 21 Rue des Lavandières, 11160 Caunes Minervois, France • fieldsres/at/gmail.com

Donald D. Hoffman • University of California, Irvine, CA, USA • ddhoff/at/uci.edu

Chetan Prakash • California State University, San Bernardino, CA, USA • cprakash/at/csusb.edu

Robert Prentner • ETH Zürich, Zürich, Switzerland • robert.prentner/at/phil.gess.ethz.ch

Structured Abstract

Paper type: Conceptual.

Backgrounds: Evolutionary theory; physics.

Perspective: Second-order cybernetics.

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: To understand objects and spacetime in observer-relative evolutionary terms.

Method: 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 Heinz 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 (2009) 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” (2009; p. 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 of von Neumann (1935/1955) or 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 (object, action) pairing 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 Hoffman and Prakash (2014) and the Markov blanket formalism of Pearl (1988) as applied to biological systems by Friston (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 of Hoffman, 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. von Foerster, 1960). Such a two-agent model is shown in Fig. 1a; here two agents S_1 and S_2 exchange observations Obs_1 and Obs_2 (cf. von Foerster, 1976, p. 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.

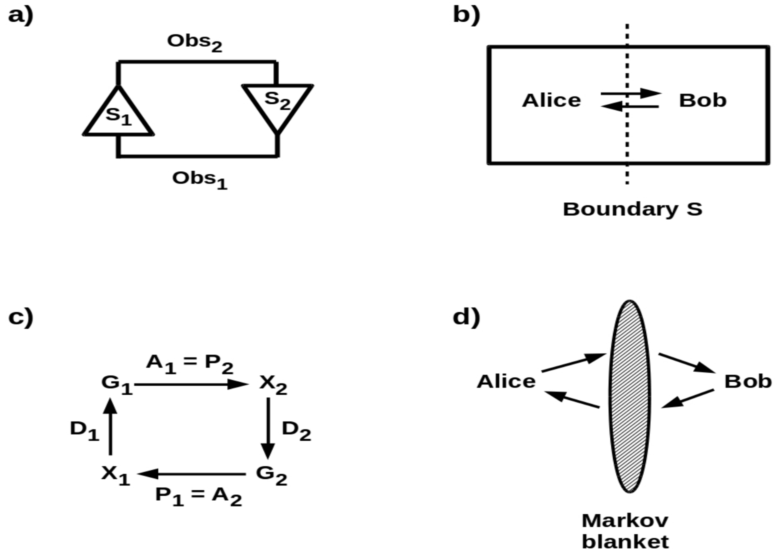

Fig. 1: Four representations of two-agent, or alternatively, agent-environment interaction. a) Two agents S_1 and S_2, here depicted as computational processes, exchange observations Obs_1 and Obs_2 (adapted from von Foerster, 1976, p. 94). b) Two agents, or alternatively two classical black boxes, Alice 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_1 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.

5. The closed-loop, two-agent exchange in Fig. 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” (von Foerster, 1960, p. 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 Fig. 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 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 2nd law is respected from each agent's individual perspective. This comports well with the probabilities that appear in the 2nd law being subjective, not objective.

6. The lack of observational access that rescues Fig. 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, 2016). 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 Fig. 1a can, therefore, also be represented as two interacting black boxes; we give them their traditional names Alice and Bob (Fig. 1B; cf. the similar construction of Glanville, 1982, Fig. 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. 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, 2016), and is hence what allows the two-agent representation in Fig. 1a.

7. The position of the boundary S separating Alice from Bob in Fig. 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, 2016); 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 Fig. 1c (adapted from Hoffman and Prakash, 2014, Fig. 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 Fig. 1c, just as they are in Figs. 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 Fig. 1c of the arbitrarily-movable inter-agent boundary S in Fig. 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 outward, 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 (Fields, Hoffman, Prakash and Singh, in review). 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 (Fig. 1d). Pearl (1988) shows that a Markov blanket appears whenever a random dynamical system is factored into parts (see Friston, 2013 or Friston, Levin, Sengupta & Pezzulo, 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 Fig. 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, Levin, Sengupta & Pezzulo, 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 Fig. 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 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 (ITP: 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 Husserl, 1913/2012 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). 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 and Stacey, 2014, p. 6). 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. 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 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 Fig. 2. In the original environment-induced decoherence process of 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 (Fig. 2a; cf. Tegmark, 2012, Fig. 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 Ollivier, Poulin and Zurek (2004; 2005), however, observers typically interact with systems of interest only via an apparatus or an ambient field such as the photon field (Fig. 2b; cf. Ollivier, Poulin and Zurek, 2005, Fig. 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.

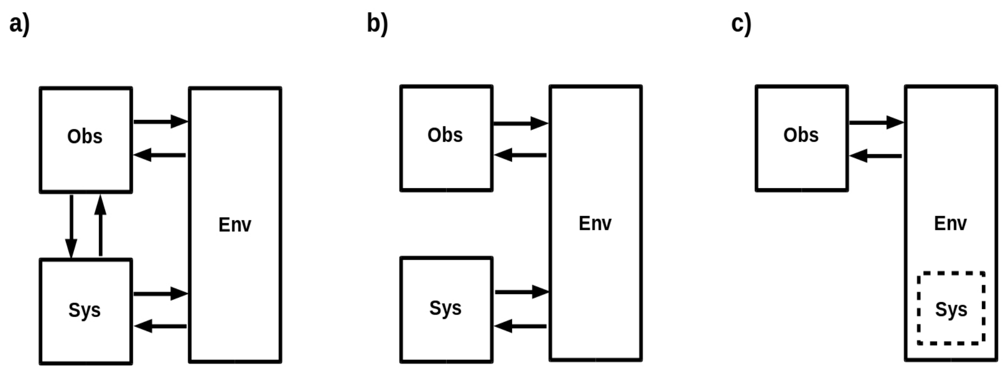

Fig. 2: Three views of decoherence. a) An observer (Obs) prepares and measures a quantum system (Sys). Both independently interact with a large surrounding environment (Env), which renders their states effectively classical by a decoherence mechanism (e.g. ambient photon scattering). b) The "environment as witness" formulation of Ollivier, Poulin and 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.

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 (Fig. 2c; cf. Fields, 2016, Fig. 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 Fig. 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 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 x_i as encoded on the interface itself, as shown in Fig. 3.

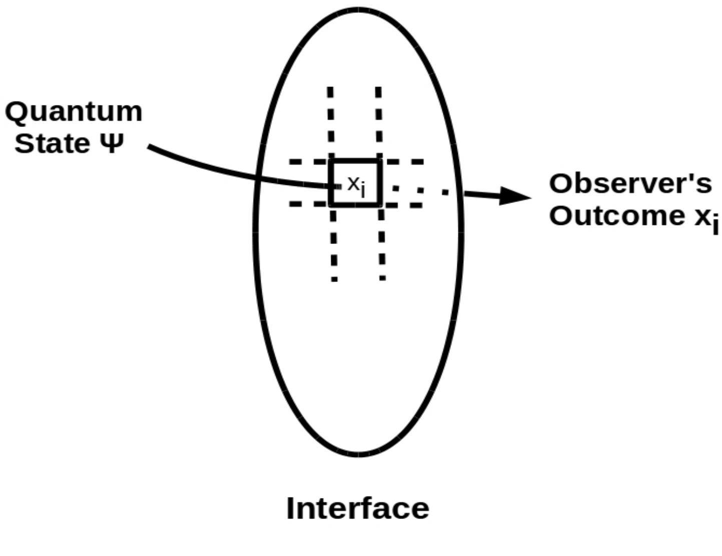

Fig. 3: Decoherence encodes a classical outcome value x_i on the observer-environment interface. Such an encoding is required if receipt of the observational outcome is to be considered to have any effect on the observer's subsequent behavior. This encoding is holographic, i.e. the only information about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

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 't Hooft (1993) and 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 n_i 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 n_i over all the degrees of freedom in $\{\xi_i\}$:

$$d_{boundary} = \Sigma_{\{\xi i\}}\, n_i.$$

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:

$$d_{bulk} = \Sigma_{\{\zeta j\}}\, m_j.$$

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, 2016). The loop from Fig. 3 back to Fig. 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, p. 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.

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 Bateson (1987) put it, “what we mean by information – the elementary unit of information – *is a difference which makes a difference*” (p. 460; emphasis in original). All of the information that agents possess is information that has had some effect on them; it is all “pragmatic information” in Roederer's (2005) sense, information that enables doing something. von Foerster (1970) makes a similar point, quoting J. Konorski: “information and its utilization are inseparable … one single process” (p. 46).

22. In the CA model of Hoffman and Prakash (2014), the recursive loop is perceive-decide-act (P-D-A) as shown in Fig. 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 Fig. 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.”

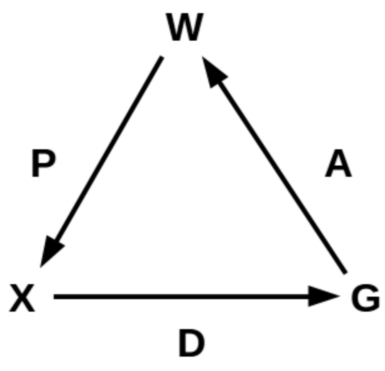

Fig. 4: A CA as defined by Hoffman and Prakash (2014) as a perceive-decide-act (P-D-A) loop through a "world" W, which takes the place of the "second agent" X_2-D_2-G_2 in Fig. 1c.

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 × G × W → Non-negative Reals. "Continued recursion" is "viability" in 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 Fig. 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 have 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:

$$EF(x|g,w) = \Sigma_{g'w'} F(x,g',w') \, Prob(g'|x,g) \, Prob(w'|g,w)$$

or making the operator actions explicit:

$$EF(x|g,w) = \Sigma_{g'w'} F(x,g',w') \, D(x,g;g') \, A(g,w;w').$$

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 of 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 (Prakash, Stephens, Hoffman, Singh & Fields, in review). The “fitness beats truth” theorem provides a formal justification for von Glasersfeld's (1981) remark that “we must never say that our knowledge is 'true' in the sense that it reflects an ontologically real world” (p. 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 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 (Fields, Hoffman, Prakash & Singh, in review). 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 and 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:

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 Fig. 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.

Fig. 5: Spatially encoding an icon allows its "parts" to each contribute to its message.

30. 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.

31. 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 Fig. 6. Here flipping a bit is “traveling” in a “direction” on a graph. The bits are independent, so the directions are orthogonal.

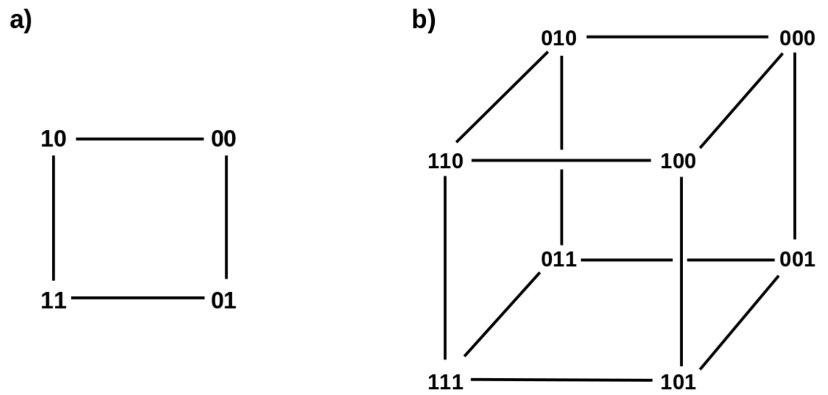

Fig. 6: Binary codes with redundancies of (a) two and (b) three. Each line represents one bit flip and hence a Hamming distance of one. The codes “00” and “11” are a Hamming distance of two apart, while “000” and “111” are a Hamming distance of three apart. Receiving “10” or “01” provides no information about the intended message, while receiving “110”, “101” or “011” suggests “111” and receiving “100”, “010” or “001” suggests “000”. Hence the three-bit code provides error correction while the two-bit code does not.

32. As can be seen in Fig. 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.

33. 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 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).

34. 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, for 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.

35. 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 and Prakash, 2015; Prakash and Hoffman, in review). 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.

36. 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 Fig. 5, that occupies space and redundantly encodes persistence.

37. Mammalian visual (e.g. Goodale and Milner, 1992) and auditory (e.g. Hickok and 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, Yoshida, Harlow & Preskill, 2015; D'Ariano & Perinotti, 2017).

Conclusion

38. In his paper introducing the “it from bit” concept, J. A. Wheeler (1990) insisted that “what we call existence is an information-theoretic entity” (p. 8), later quoting 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” (all p. 10). von Foerster could well have added: spacetime is the eigenform that by remaining constant enables actions.

39. 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.”

40. 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.

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.
Bateson G. (1987) Steps to an Ecology of Mind. Jason Aronson, London.
Baillargeon R. (2008) Innate ideas revisited: For a principle of persistence in infants' physical reasoning. Perspectives on Psychological Science 3(1): 2-13.
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.
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.
Bousso R. (2002) The holographic principle. Reviews of Modern Physics 74(3): 825-874.
Conway J. & Kochen S. (2006) The free will theorem. Foundations of Physics 36(10): 1441-1473.
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, U.K.: 38-43.
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. (2016) Building the observer into the system: Toward a realistic description of human interaction with the world. Systems 4: 32.
Foerster H. von (1960) On self-organizing systems and their environments. In: M.C. Yovits M. C. & Cameron S. (eds.) Self-Organizing Systems. Pergamon, London: 31-50. Reprinted in Foerster (2003): 1-19.
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.
Foerster H. von (1973) On constructing a reality. In: Preiser F. E. (ed.) Environmental Design Research, Vol. II. Dowden, Hutchinson & Ross, Stroudsburg, PA: 35-46. Reprinted in Foerster (2003): 211-227.
Foerster H. von (1976) Objects: Tokens for (eigen-) behaviors. ASC Cybernetics Forum 8(3-4): 91-96. Reprinted in Foerster (2003), pp. 261-271.
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.

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. Preprint arxiv:1003.5209v1 [quant-ph].
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-Miffin, Boston.
Glanville, R. (1982) Inside every white box there are two black boxes trying to get out. Behavioral Science 27(1): 1-11.
Glasersfeld E. von (1981) The concepts of adaptation and viability in a radical constructivist theory of knowledge. In: Sigel I. E., Brodzinsky D. M. & Golinkoff R. M. (eds.) Piagetian Theory and Research. Erlbaum, Hillsdale, NJ: 87-95.
Goodale M. A. & Milner A. D. (1992) Separate visual pathways for perception and action. Trends in Neurosciences 15(1): 20-25.
Hensen B. et al. (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. and 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.
Hooft G. 't (1993) Dimensional reduction in quantum gravity. In: Ali A., Ellis J. & Randjbar-Daemi S. (eds.) Salamfestschrift. World Scientific, Singapore: 284-296.
Husserl E. (1913/2012) Ideas: General Introduction to Pure Phenomenology. Routledge, London (Originally published in German, 1913).
Kauffman L. H. (2003) Eigenforms – Objects as tokens for eigenbehaviors. Cybernetics and Human Knowing 10(3-4): 73-90.
Kauffman L. H. (2009) Reflexivity and eigenform: The shape of process. Constructivist Foundations 4(3): 121-137.
Kauffman L. H. (2011) Eigenforms and quantum physics. Cybernetics and Human Knowing 18(3-4): 111-121.
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.
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., and 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.
Moore E. F. (1956) Gedankenexperiments on sequential machines. In: Shannon C. W. & McCarthy J. (eds.) Automata Studies. Princeton University Press, Princeton, NJ: 129-155.
Neumann J. von (1935/1955) Mathematical Foundations of Quantum Mechanics. Princeton University Press: Princeton, NJ (Originally published in German, 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.
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.
Roederer J. (2005) Information and its Role in Nature. Springer, Berlin.
Rubino G. et al. (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.
Spelke E.S. (1994) Initial knowledge: Six suggestions. *Cognition* 50(1-3): 431-445.
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.
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, U.K.
Trivers R. L. (2011) The Folly of Fools. Basic Books, New York.
Wallace D. (2008) Philosophy of quantum mechanics. In: Rickles D. (ed.) The Ashgate Companion to Contemporary Philosophy of Physics. Ashgate, Aldershot, UK: 16-98.
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.
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.

The authors

Chris Fields (Ph.D. 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 (Ph.D. 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* (W.W. Norton, 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. Prof. 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; this is the subject of his TED Talk entitled “Do we see reality as it is?” and of an article in *The Atlantic* entitled “The case against reality.” It has also led to a “conscious realism” theory of consciousness – which proposes a formal model of consciousness and a new solution to the mind-body problem.

Chetan Prakash (Ph.D. Mathematical Physics, Cornell University, 1982) has published, with Bruce Bennett and Don Hoffman, the book *Observer Theory*, with Don Hoffman the seminal paper “*Objects of Consciousness*” and, with Hoffman, Stephens, Singh and Fields “*Fitness Beats Truth in the Evolution of Perception*” (in review) and, with Fields, Hoffman and Singh, “*Conscious Agent Networks: Formal Analysis and Application to Cognition*” (in review). 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 lead us to perceive an “objective” reality with ever-increasing accuracy. Dr. Prakash is also a senior instructor in Aikido, ranked 6th degree black belt, and has practiced Aikido for 33 years.

Robert Prentner (Ph.D. Physical Chemistry, ETH Zürich, 2013) has been a visiting scholar at Stanford University‘s Center for the Explanation of Consciousness. Since Fall 2013 he is 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."