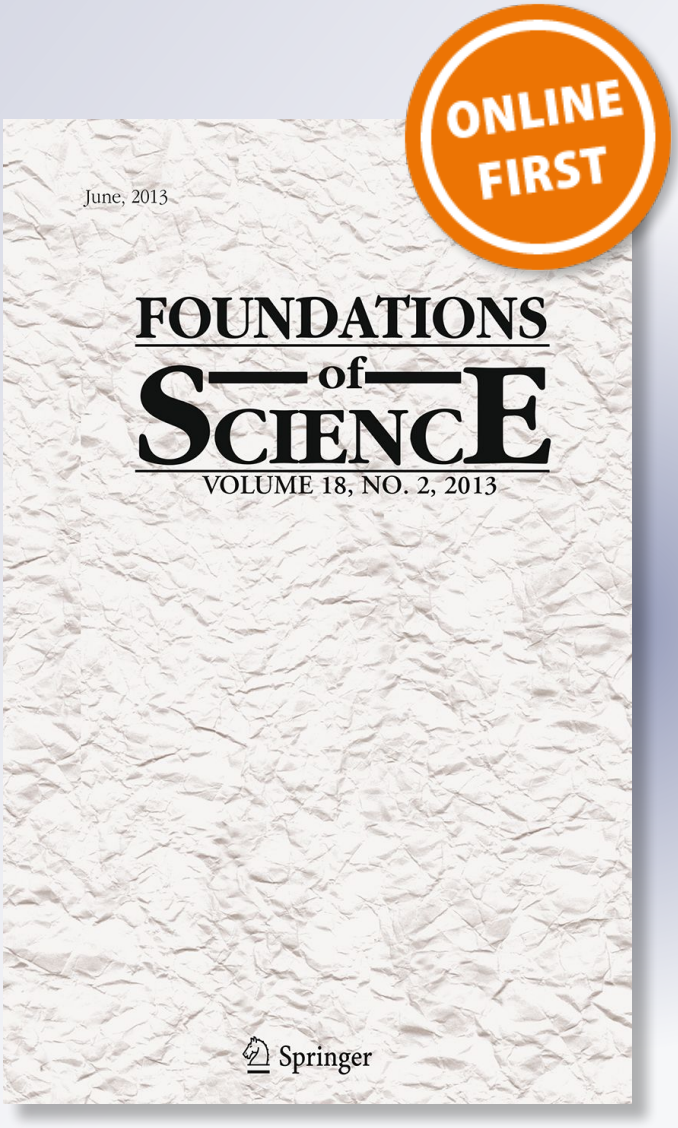
On Invention of Structure in the World: Interfaces and Conscious Agents

Chetan Prakash

Foundations of Science
The official Journal of the Association for Foundations of Science, Language and Cognition

ISSN 1233-1821

Found Sci
DOI 10.1007/s10699-019-09579-7



Springer

Your article is protected by copyright and all rights are held exclusively by Springer Nature B.V.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Author's personal copy

Foundations of Science
https://doi.org/10.1007/s10699-019-09579-7

On Invention of Structure in the World: Interfaces and Conscious Agents

Chetan Prakash[1]

© Springer Nature B.V. 2019

Abstract
The *Interface Theory of Perception*, as stated by D. Hoffman, says that perceptual experiences do not to approximate properties of an "objective" world; instead, they have evolved to provide a simplified, species-specific, user interface to the world. *Conscious Realism* states that the objective world consists of 'conscious agents' and their experiences. Under these two theses, consciousness creates all objects and properties of the physical world: the problem of explaining this process reverses the mind-body problem. In support of the interface theory I propose that our perceptions have evolved, not to report the truth, but to guide adaptive behaviors. Using evolutionary game theory, I state a theorem asserting that perceptual strategies that see the truth will, under natural selection, be driven to extinction by perceptual strategies of equal complexity but tuned instead to fitness. I then give a minimal mathematical definition of the essential elements of a "conscious agent." Under the conscious realism thesis, this leads to a non-dualistic, dynamical theory of conscious process in which both observer and observed have the same mathematical structure. The dynamics raises the possibility of emergence of combinations of conscious agents, in whose experiences those of the component agents are entangled. In support of conscious realism, I discuss two more theorems showing that a conscious agent can consistently see geometric and probabilistic structures of space that are not necessarily in the world *per se* but are properties of the conscious agent itself. The world simply has to be amenable to such a construction on the part of the agent; and different agents may construct different (even incompatible) structures as seeming to belong to the world. This again supports the idea that any true structure of the world is likely quite different from what we see. I conclude by observing that these theorems suggest the need for a new theory which resolves the reverse mind-body problem, a good candidate for which is conscious agent theory.

Keywords Mind-body problem · Evolutionary games · Conscious agent · Consciousness · Physicalism

✉ Chetan Prakash
cprakashresearch@gmail.com

[1] CSUSB, 22785 Raven Way, Grand Terrace, San Bernardino, CA 92313, USA

Published online: 24 January 2019

Author's personal copy

1 Introduction

The prevalent view of perception is *physicalist*: perception is conceived as a reconstruction of aspects of an objective (physical) world, i.e., one whose structure is independent of any acts of observation. As such, perception is considered to be, by and large, approximately veridical; moreover it is believed that evolution drives perceptual systems to ever-greater veridicality. This view is standard in cognitive science, e.g., Marr (1982), Palmer (1999), Giesler (2003), Pizlo et al. (2014) (though it is not held exclusively: e.g., Chemero 2009). Physicalism is a view that is also held by the vast majority of researchers in the physical and biological sciences. However, there is much evidence to question the assumption that perception is veridical and that evolution drives perception to become more so: see, e.g., Koenderink (2013), Koenderink (2015), Hoffman (2000), Mausfeld (2015), Hoffman (2019).

An extension to the physicalist viewpoint is the idea that consciousness is an emergent property of highly complex (though hitherto unspecified) neuro-biological processes. Chalmers (1995) referred to the problem of specifying the process of this emergence as the "hard problem" of consciousness.

This article challenges these physicalist assumptions in two ways. Firstly, assuming the existence of a fixed, observer-independent objective reality, I state a theorem which achieves a kind of "reductio ad absurdum" to the claim of veridicality of perception. Specifically, the theorem, utilising standard techniques of Bayesian analysis and evolutionary game theory, says that resource-acquisition strategies relying on true perception are likely to be driven to extinction by strategies relying instead on fitness. The theorem moreover gives a lower bound on the likelihood of said extinction: this bound goes to 1 as the size of the organism's perceptual space increases to infinity. This strongly challenges the idea that natural selection leads to more accurate percepts. A question then arises: if we are not attuned to accuracy in our perceptions but only to those which are fitter, what then are we perceiving of the world?

This result motivates a discussion of the interface theory of perception (Hoffman 2009; Hoffman et al. 2015), which recognises that perceptions do not necessarily represent an objective reality but have evolved to lie in an *interface* with reality; one that may bear little or no resemblance to that reality. With this understanding we are prepared to define a "conscious agent," or CA, as an entity with a mathematically precise structure (Hoffman and Prakash 2014). CAs are descriptions of the loop of perception-decision-action and are meant to represent conscious processes: the *Conscious Agent Thesis* states that *any* conscious process may be described by a CA. Moreover, a study of CA networks (Fields et al. 2017) shows that collections of interacting CAs can be thought of as single CAs; in fact, the world of any given CA can be construed as another CA.

The second challenge to the physicalist view then arises in the context of this CA structure. I present two theorems about the "invention" of structure in the world. The first theorem states that the CA's view of the world may include geometrical structures (such as 3-dimensional space) that need not actually be a structure in the world. The second theorem states that the CA's view of the world may include probabilistic structures (such as its measurable space) that need not be a structure in the world.

A logical consequence of physicalist theories is the belief that physical objects and interactions, in that they are assumed to have an objective existence, have *causal* powers. In particular, brain activity causes consciousness. Our development in what follows will strongly challenge these "obvious" beliefs.

Author's personal copy

We conclude by asking some questions about our approach to understanding consciousness and its relation with the physical world and speculating on a way of finding an answer.

2 Interface Evolutionary Theories

In this section we will assume, at first, the existence of an objective, observer-independent world that has a given, unchanging underlying structure; and that said structure can be given mathematical representation with arbitrary accuracy. This is a foundational belief in a physics that can imagine a "theory of everything." We call `veridical perceptions` those which are attuned to the structure of the world.

We also assume that our perceptual capacity has evolved as a *user interface* between us and our world, but we wish to question the assumption that fitness necessarily implies veridicality of perception. So we introduce the notion of `interface perceptions`: those that are attuned only to evolutionary fitness for a given species, in its environment and performing actions from within some given class. The space of possible such perceptions is called the `interface` for that species. It is indeed a *user* interface, in that it supplies the model for perceptions and actions best qualified for the survival of the species. Objects that are reified in this interface and interactions that seem to take place between these objects ("seem to," because they are taking place within the interface) constitute the "physics" of that species. Physicalism states that interface perceptions are indeed veridical perceptions—at least for the human species—but we make no such assumption: we will see that, in our model at least, they differ dramatically.

In modeling perceptual strategies and in working with conscious agents, we will employ mathematics from three areas: probability theory (including Bayesian analysis), evolutionary game theory and geometry. Probability theory is a foundational theory of inference under uncertainty and is therefore at the basis of most cognitive modeling. Bayesian analysis is a method of inference: it allows optimal updating of the estimated state of a system from its perceptions. Bayesian decision theory, a standard tool in conventional theories of perception, is of wide applicability in understanding inference. Evolutionary game theory (e.g., Novak 2006) has been described in Dennett (1995) as an ontologically neutral "`universal Darwinism`": 'Whether evolution has a purpose or not, it is the mechanism for the adaptation of a species to its world.' And geometry is the study of space by means of its symmetries: space is the underlying geometrical structure, if any, of the interface of a conscious agent.

2.1 Modeling Perceptual Strategies and Fitness

We assume that the states of the world can be described by a set W and the states of the perceptual space by a set X. Each of these comes equipped with a `measurable structure`, i.e., a collection of `measurable sets` which includes the whole set and which is closed under countable union and complementation. This is the minimal structure, via the Kolmogorov axioms, within which we can assign probabilities. In addition, we assume that there is an *a priori* probability measure μ on the world state space W.

An `ideal perceptual map` is a measurable function p from W to X, i.e., a function that assigns, to each world state $w \in W$, a perception $x = p(w)$ in X. More generally we wish to allow for dipersion, or uncertainty in the percept, so we define a general `perceptual map` as a measurable Markovian kernel p from $W \times X$ to X, i.e., a function

Author's personal copy

that assigns, to each world state $w \in W$, a probability measure on the perception space: $p(w, \mathrm{d}x) \in X$.[1] In this article we will, for the sake of simplicity, restrict attention to ideal perceptual maps.

Given an organism, environment and action class, to each state w of the world will correspond a `fitness`, measured as a non-negative number. This `fitness landscape` is then a function $f : W \to [0, \infty]$.

2.2 The Resource-Strategy Model: "Veridical" and "Interface" Strategies

An `evolutionary resource game` is one in which two organisms employ their individual strategies to compete in acquiring available territories containing desired resources. One player estimates—and so chooses—its best available territory and then the other chooses its own best territory from among those still available. The object is to maximize fitness as the payoff. The states of the world, then, represent territories containing different quantities of a resource and the player chooses that territory that is optimal by a criterion specific to that player. We call such a map a `perceptual strategy` and identify two possible perceptual strategies for the players in our game:

"Veridical" strategy For each of its sensory states x, the player first computes the territory w_x that is most likely to have given rise to that sensory state, i.e., the maximum *a posteriori* (MAP) estimate computed from the Bayesian posterior distribution. The strategy then picks that territory w_{x_0} that has maximum fitness among these estimates:

$$f(w_{x_0}) = \max\{f(w_x) | x \in X\}$$

"Interface" strategy The player computes the `expected fitness` $F(w)$ of each territory w, by averaging, over the Bayesian posterior distribution, the fitness values of world states corresponding to that territory. The player then chooses the territory with maximum expected fitness.

It is assumed here that there *is* an objective world of territories available to both players. The Interface player, however, does not concern itself in any way on the most likely world state and is only interested in expected fitness.

In order to perform the Bayesian computations, we assume further structure in W, which for our present purposes may be assumed to be a rectangular finite-volume subset of n-dimensional Euclidean space, or even some finite set. These are the instances most common in experimental psychology. In either case there is a uniform probability measure on the world states, which we will write as $\mathrm{d}w$; we will assume that the *a priori* measure μ on W has a continuous *density* with respect to the uniform measure: $\mu(\mathrm{d}w) = g(w)\mathrm{d}w$, for some function g (assumed continuous in the non-discrete case). This then allows us to write a Bayes' formula for the *a posteriori* probability that a world state w could have given rise to the perception x: if we write the `likelihood function` as $\mathbb{L}(x|w)$, then the a posteriori probability is

$$\mathbb{P}(w|x) = \frac{\mathbb{L}(x|w)g(w)}{\sum_{w' \in W} \mathbb{L}(x|w')g(w')}$$

[1] If the space X is discrete, this just means that each pair (w, x) is assigned a probability $K(w, x)$ between 0 and 1 and that $\sum_{x \in X} K(w, x) = 1$, for all $w \in W$. The ideal strategy is then just a "Dirac" kernel: $K(w, \mathrm{d}x) = \delta_{p(w)}(\mathrm{d}x)$, for some function $p : W \to X$; the right-hand side is the Dirac point measure assigning the value 1 to any measurable set containing $p(w)$ and zero otherwise. We will generalize this further later in this article, when we allow the current perceptual strategy to depend also on the previous percept.

Author's personal copy

Now suppose we are given an ideal perceptual map $p : W \to X$. Then the likelihood function is $\mathbb{L}(x|w) = 1_{p^{-1}(x)}(w)$, i.e., the indicator function of the `fibre of p over x`. This is the subset of world states that *could* have given rise to x as a perception. So we can rewrite our equation as

$$\mathbb{P}(w|x) = \frac{1_{p^{-1}(x)}(w)g(w)}{\sum_{w' \in p^{-1}(x)} g(w')}$$

This then allows us to compute both the MAP estimate for the veridical strategy, as well as the expected fitness for the interface strategy. Note that in maximizing $\mathbb{P}(w|x)$, we need only maximize $g(w)$ over the fibre:

MAP estimate for percept x: w_x is an element of the fibre $p^{-1}(x)$ that maximizes $g(w)$. If there is more than one maximizer, we include all such as MAP estimates.

Expected fitness of percept x: This is the average of the fitness function, over the a posteriori probability distribution, in the fibre over x:

$$\langle x \rangle_p := \sum_{p^{-1}(x)} f(w)\mathbb{P}(w|x) = \frac{\sum_{p^{-1}(x)} f(w)g(w)}{\sum_{p^{-1}(x)} g(w')}$$

Note that, for both strategies, Bayes' theorem is essential.[2]

Games with these strategies were first studied studied via simulations in Mark et al. (2010) and theoretically in Prakash et al. (2018) and references therein. The strategies were there referred to as "truth" and "fitness" respectively.

2.3 Fitness Beats Truth

Imagine an evolutionary game where we pit the Interface and Veridical strategies against each other. Upon computing a payoff matrix and utilizing basic theorems of evolutionary game theory, we obtain the following theorem that shows that interface strategies will generally drive veridical ones to extinction (proved in Prakash et al. 2018):

Theorem 1 (Fitness Beats Truth) *Generically over all possible fitness functions and a priori measures, the probability that the Interface perceptual strategy strictly dominates the Veridical strategy is at least* $\frac{|X|-3}{|X|-1}$, *where* $|X|$ *is the size of the perceptual space.*

So as the size $|X|$ of the perceptual space increases, the generic probability that Interface dominates Veridical becomes arbitrarily close to 1: in the limit as $|X| \to \infty$, Interface will generically strictly dominate Veridical, so driving the latter to extinction. As mentioned in the introduction, this means that our perceptions have evolved to ensure our survival, not to depict whatever reality is "out there."

This result suggests that our physical vocabulary, regarded as an extension of our *perceptual* predicate vocabulary, is the wrong vocabulary for describing the causal structure of the real world.

[2] The above can easily be generalized (Prakash et al. 2018) to the situation where the perceptual map is a Markov kernel $p(w, x)$: replace $1_{p^{-1}(x)}$ by $p(w; x)$ and the sums by sums (or integrals) over the whole of W.

Author's personal copy

3 Conscious Agents

The Fitness Beats Truth result suggests the need for a new theory of perception. In the Interface Theory, proposed by Don Hoffman in Hoffman (2009), space-time is thought of as a labelling, or *coordinatization*, of certain aspects of our interface, analogous to the way pixel position gives coordinates to a computer desktop. Physical *objects* are then icons on that "desktop." These physical objects-icons seem to interact via "physical" law, but physical causality itself is a fiction: everything *experienced* is happening on the desktop and what is happening in the innards of the "computer," i.e., the objective world, may bear no relation to what is experienced, or to what is inferred from what is experienced. Thus we should not take our perceptions *literally*, but this does not mean that we should not take them *seriously*: would I let (the icon of) my body, e.g., step in front of (the icon of) a train?

We have concluded that evolution favors Interfaces as perceptual strategies and in the next section we will see that structural properties of the world, such as geometrical space and probability assignments, also appear to be species-dependent constructions. It seems, then, that consciousness is primary: *perceived* structure in the world is secondary, arising in a species-adapted manner. Moreover, it is well-known that our current science has been compllletely unable to demonstrate the emergence of qualia and consciousness from brains and their bodies. This invites us to consider developing a new model of consciousness, one in which consciousness is primary and the "physical" world emergent. Such a model, proposed in Hoffman and Prakash (2014), goes as follows.

A conscious agent interacts with its world. What is an agent? We propose that the core aspects of conscious agency can be modeled as a 7-tuple, as follows. Let the states of the world, the perceptions of the agent and the actions of the agent on the world be probability spaces W, X and, G, respectively, and let N be an integer that counts the number of perception-action events after some arbitrary start. The act of perception is represented by a Markov kernel $P : W \times X \to X$. This means that the probability that the current perception will be x' depends on the previous perception x and the current world state w: we denote it as $P(w, x; \mathrm{d}x')$. (Note that in the discreet case this just means that the probability is $P(w, x; x')$, where, for any pairing of world state $w \in W$ and perception state $x \in X$, we have $\sum_{x' \in X} P(w, x; x') = 1$.) Having had a perception, a conscious agent makes a decision to act. This decision is also represented by a Markov kernel, $D : X \times G \to G$ as follows: given the previous action g and the current perception x, the probability that the current action will be g' is $D(x, g; \mathrm{d}g')$. Finally, the action of of the agent on the world is represented by a third Markov kernel $A : G \times W \to W$: given the previous world state w and the current action state g, the probability that the current world state will be w' is $A(g, w; \mathrm{d}w')$. We define a `Conscious Agent`, denoted `CA`, to be the 7-tuple $\langle W, X, G, P, D, A, N \rangle$ as above.[3]

[3] Prior to this definition, given in Fields et al. (2017), a CA was defined in Hoffman and Prakash (2014) more restrictively as what should now be termed a `forgetful CA`: we take for the perception kernel a Markov kernel $P : W \to X$, where the probability that the new perception is x depends on the current world state w: it is denoted $P(w; \mathrm{d}x)$. The decision kernel is a Markov kernel, $D : X \to G$. Here, given the current perception x, the probability that the next action will be g is $D(x; \mathrm{d}g)$. Finally, the action Markov kernel is $A : G \to W$. This means that, given the current action state g, the probability that the next world state will be w is $A(g; \mathrm{d}w)$.

Author's personal copy

We propose that this is the minimal structure common to all conscious agents.[4] In modeling the causal loop of perception-decision-action, the use of Markov kernels as against punctual functions allows us to model the theorist's uncertainty in perception, decision and action and, in particular, the probabilistic nature of the decision kernel allows us to model free-will. Indeed, we propose the

Conscious Agent Thesis *Any process of consciousness can be modeled by a 7-tuple* $\langle W, X, G, P, D, A, N\rangle$ *as above.*

Any precisely describable conscious process can be subjected to the test of fitting its description to this model. In other words, given a specific conscious process, we can ask: are there spaces of perceptual states, actions and world states that encapsulate the states and actions involved in this process? Are there functional and causal relationships from world states to perception states, from perceptions to actions and from actions to effects on the world? If there is dispersion involved, i.e., uncertainty in the functional relationships, can they be sufficiently described by assigning probabilities (and, therefore, implicitly assigning measurable structure) to the three kinds of state?[5] And, if so, do the functional relationships identified above have the structure of Markovian kernels, that assign to each pair of source states, as in the definition of CA above, a probability measure on the set of target states? Thus we see that, in principle, the Conscious Agent thesis is falsifiable, and is therefore a "scientific" thesis in conventional terms.

We note that a model very similar to that proposed in Hoffman and Prakash (2014) was given by Ay and Löhr (2015), who present a mathematical model of von Uexkül's "sensorimoter loop" (Uexküll 2014). They define a Markov process involving four sequence spaces and four time-dependent Markov kernels, plus a discrete time counter. They show that their "sensory" σ-algebra is a subalgebra of the σ-algebra on the world states. In our model, this is not the case because we have not seperated sensory from perceptual states. Our P does, however, *induce* a σ-algebra on W (see Sect. 4.2 below), which is, by the measurability of P, a subalgebra. Since we do not assume veridicality of perceptions (as, it seems, Ay and Löhr do), their sensory state space would then be, in our model, a part of the world.

4 Invention of Space and Probabilistic Structure

With the mathematical structure so introduced, we can explore the question of the veridicality of our perceptions of the world and of the accuracy of our perceived actions on the world.

For simplicity we will restrict ourselves to the simplest instance of punctual conscious agents:

Assumption All conscious agents under consideration will be `punctual conscious agents`, namely ones where all the Markov kernels are Dirac kernels, so that they are representable simply as functions. We will think of P as a function from $W \times X$ to X, D as a function from $X \times G$ to G and A as a function from $G \times W$ to W.

[4] Here, by "conscious agent" we are not restricting our attention to *human* agents, self-conscious and aware of their perceptions, decision and actions, or even to human agents in *any* state of consciousness or lack thereof.

[5] In the non-dispersive instance, it suffices, for the sake of consistency in the definition of conscious agent, to put the discrete σ-algebra on the spaces.

Author's personal copy

4.1 Physical Space

We will treat the concept of "space" as a geometry in the Kleinian sense: a `geometrical space` is a set S of points together with a group H acting on S transitively and faithfully.[6] We summarize this by saying that S is an H-`space` and we will say that S is `geometrized` by H.

There can be further structure, such as a metric on S that measures the degree of separation between points or, further, S could be an n-dimensional differentiable manifold acted on by a Lie group H. The group is called a `symmetry group` because it leaves invariant certain equivalence classes of figures in, i.e., subsets of, S. As examples we have the Euclidean group of classical geometry acting on 3-space, which preserves distances between points (and leaves invariant, e.g., the set of all triangles congruent to a given one); the affine group in 3 dimensions preserving parallelism (and so leaving invariant the set of parallelograms); the projective group on $\mathbb{R}^3$, preserving cross-ratio (and leaving invariant the set of quadrilaterals); or the Lorentz group of special relativity acting on real 4-space, preserving Minkowski distance.

Roughly speaking, the Invention of Space theorem below says that, if a conscious agent's actions form a group G, if this group G has a transitive action on the perception space and if that agent's perceptions are "tuned" to its actions, then the agent's perception space can be partitioned so that the set of these partitions is a G-space for that group. If this is so, the agent will be able to *geometrize* its interface by that G-space.

It is important to note that, although the agent's action space G consists of actions on the world states W, then even though G is assumed to be a group, and even though this group acts (abstractly) on X, the said action of G on the world will not, in general, be a *group action*: it will not be transitive (since the world is expected to be (much) larger than G) and there is no reason to suppose that the effects of the CA on the world are reversible, as would be required by a group action.[7]

In the simplest instance, that of a `pure-space conscious agent`, we suppose the entire perceptual space X is a G-set, where the action space forms the group G. By definition, for a current percept x_1 from the current world state w_1, and for any current action g_1 in G, the new world state is $w_2 = A(g_1, w_1)$. Then the new percept will be $x_2 = P(w_2, x_1)$.

Definition 1 We say that an agent's perceptions and actions are `attuned` if, for any current percept x_1 from the current world state w_1; and for any current action g_1 in G, we have for the *next* percept that $x_2 = g \cdot x_1$.

This means that the next percept is precisely that which would have obtained had the agent's action g acted on the *percept* x, even though it actually acted on W.

[6] The group H `acts` on S if there is a mapping $H \times S \ni (h, s) \times S \mapsto h \cdot s \in H$, such that the identity $e \in H$ acts as $e \cdot s = s, \forall s \in S$ and whenever $h, k \in H$ and $s \in S$, $h(k \cdot s) = (hk) \cdot s$. In the last equality, hk is a product in the group H, while $k \cdot s$ and $(hk) \cdot s$ express the action of the group elements k and hk, respectively. The action is `transitive` if, for all pairs of elements of S, there is a group element taking one to the other; it is faithful if there is no $s \in S$ with $h \cdot s = s, \forall h \in H$.

[7] Let $w' = A(g, w)$ and $w'' = A(g^{-1}, w')$. Then $w'' = A(g^{-1}, A(g, w))$. If this were a *group* action, we would require $w'' = A(g^{-1}g, w)) = A(\iota, w) = w$, where ι is the identity of G.

Author's personal copy

Definition 2 The `conditional perception mapping at` x, or `cpm`$_x$, is $p_x(\cdot) : W \to X$, for $x \in X$, given by $W \ni w \mapsto p_x(w) := P(w,x)$. The `cpm fibre space of P at x` is the collection of subsets of W given by $F_x = \{p_x^{-1}(x') | \, x' \in X\}$.[8]

Fix $x \in X$. Then the group action of G on X induces an action of G on F_x, as follows: using the same notation for this action, we may define $g \cdot p_x^{-1}(x')$ by $p_x^{-1}(g \cdot x')$. Attunement then means that the fibre over the new percept x_2, namely $p_{x_1}^{-1}(x_2)$ is none other than the fibre over the previous percept x_1, now acted upon by g: namely, $g \cdot p_x^{-1}(x_1) = p_x^{-1}(g \cdot x_1)$. The Invention of Space theorem below asserts that the perception-action experiences of an agent attuned as above will admit G as a group of symmetries on each cpm fibre space of P, which is a structure in the world W. The agent "sees" the world as having the geometrical structure internal to it, even though the world may have no such structure but is merely *amenable* to such an "illusion" on the part of the agent.

More generally, many of the CA's perceptions may not be associated with any notion of "space," since only some subset of perceptions would be spatially located. Also, multiple percepts may be spatially located together. So let X_s be a subset of X, with a `partition` $S = \{S_i\}_{i \in \mathcal{I}}$ of X_s, i.e., the S_i, for $i \in \mathcal{I}$ an indexing set, are disjoint subsets whose union $\cup_{i \in \mathcal{I}} S_i$ is X_s. Let $\pi : X_s \to S$ be the `natural map` that assigns to percepts in X_s the subset S_i that they belong to. Then we will say that the CA's `perceptions and its actions are attuned over` S, if the set G of actions on the world is a group, the set S is a G-set and for any current percept $x_1 \in S_i$ from the current world state w_1 and current action g_1 in G (so that new world state is $w_2 = A(g_1, w_1)$), we have for the next percept x_2 that $x_2 = P(w_2, x_1) \in g_1 \cdot S_i$. When this holds, it would be reasonable to call S the `spatial set` of X. The following theorem then generalizes one announced in Hoffman et al. (2015):

Theorem 2 (Invention of Space) *Suppose that a (punctual) conscious agent's action space G has the structure of a group, there is a spacial set S associated to X as above and the agent's perceptions and its actions are attuned over S. Then the perception-action experiences of this agent will admit G as a group of symmetries on the fibres of $\pi \circ p_x$, for each $x \in X_S$, in the world W.*

4.2 Measurable Structures

Suppose we are given measurable sets $\bar{X}$, and $\bar{G}$ and a measurable function $\bar{D}$ from $\bar{X} \times \bar{G}$ to $\bar{G}$. Suppose $\bar{W}$ is another set, $\bar{P}$ is *any* function from $\bar{W} \times \bar{X}$ to $\bar{X}$ and $\bar{A}$ *any* function from $\bar{G} \times \bar{W}$ to $\bar{W}$. Note that the only difference from the definition of a CA is that we are not asserting a measurable structure on W and therefore we are not asserting measurability of P and of A. $\left\langle \bar{W}, \bar{X}, \bar{G}, \bar{P}, \bar{D}, \bar{A}, \bar{N} \right\rangle$ is called a `reduced conscious agent` or RCA (introduced in Fields et al. 2017). An RCA represents that which is internal to the conscious agent, together with its interface.

Now specialize to a *forgetful*, punctual RCA, i.e., one for which the functions are independent of the prior states represented by the second argument above and e.g., $\bar{P}$ is *any* function from $\bar{W}$ to $\bar{X}$ etc. Then the function $\bar{P}$ automatically induces a measurable structure on $\bar{W}$: if $\bar{\mathcal{X}}$ is the σ-algebra on $\bar{X}$, then it is straightforward to show that the collection

[8] By $p_x^{-1}(x')$ we mean the set of all $w \in W$ such that $p_x(w) = P(w,x) = x'$.

Author's personal copy

$\bar{\mathscr{W}} = \{P^{-1}(B) | B \in \bar{\mathscr{X}}\}$ is a σ-algebra, called the σ-`algebra induced by` P `on` $\bar{W}$. P is then a measurable function. Thus we have proved:

Theorem 3 (Invention of Probabilistic Structure) *Suppose* $\left\langle \bar{W}, \bar{X}, \bar{G}, \bar{P}, \bar{D}, \bar{A}, \bar{N} \right\rangle$ *is a (forgetful, punctual) reduced conscious agent. If* $\bar{A}$ *is a measurable function with respect to the* σ*-algebras* $\bar{G}$ *and* $\bar{\mathscr{W}}$ *(induced by* $\bar{P}$*), then* $\left\langle \bar{W}, \bar{X}, \bar{G}, \bar{P}, \bar{D}, \bar{A}, \bar{N} \right\rangle$ *is a (forgetful, punctual) conscious agent.*

Again, we see that with an attunement between perceptions and actions, here (in regards to measurability) a probabilistic one, the agent sees a structure "in the world," even if that world does not have this structure.

5 Discussion and Conclusion

One might object to Fitness Beats Truth Theorem as follows: "But don't fitness functions correspond to the truth, i.e., aren't they "homomorphic" to the objective structure of the world?" Given that fitness is species-specific, the number of possible perceptual strategies (and the payoffs they receive as a result of their fitness functions), in step with the number of species, greatly exceeds any set of structure-preserving functions. So it is highly unlikely that the collection of all possible fitness functions represents any given objective structure. Another objection goes as follows: "There is consistent agreement on things like 3-D shapes, textures etc. How can you say that these are fictions of our interface?" The Invention theorems stated above strongly militate against such intuitions, strong as they (conventionally) may be: the observer sees spatial and probabilistic structures in the world, but it is the *agent* that has this structure; the world may not have this structure!

Bertrand Russell has said: "Thus it would seem that, whenever we infer from perceptions, it is only structure that we can validly infer..." (Russel 1959). The investigations herein suggest that this statement is correct; though not as Russell intended: it is not the structure of the world, but that of our perceptions, that is inferred.

It seems, then, that space-time is a description, by *human* conscious agents, of location and dynamics on *their* perceptual interface, that the "objects" of Physics are icons on that interface and that "phenomena," as they appear to us, are properties of the dynamics of those icons on that interface.

Future enhancements of the results described above would generalize the invention theorems to non-punctual and non-forgetful CAs and include a theorem on invention, as suggested by Fields (2018), demonstrating the inability of a CA to ascribe permanence to objects in the world. Various other questions suggest themselves for further research, such as: How do conscious agents combine, statically or dynamically, to produce new conscious agents? How do we build circuits of conscious agents and explore their properties, including memory, predictive coding etc.? A start in this direction is in Fields et al. (2017).

The *combination problem* was articulated by William James (James 1895) as the problem of how qualia (James refered to "feelings") could possibly combine to produce a new, higher-level quale. The combination problem has been an issue for panpsychism, which ascribes consciousness to microphysical entities: to quote Chalmers, "[The combination problem] is roughly the question: how do the experiences of fundamental physical entities such as quarks and photons combine to yield the familiar sort of human conscious experience that we know and love" (Chalmers 2016). As has been pointed out by Angela

Author's personal copy

Mendelovici (2019), the issue is not restricted to panpsychism, but is common to all philosophical approaches to consciousness. A theory of how conscious agents may combine to produce new conscious agents would indeed provide a possible solution to the combination problem.

Following on this we may ask: how does such a combination evolve an interface? And how would space-time emerge from consciousness? Our view is that space-time emerges as an efficient coding scheme, one which allows a given agent to reduce the incredibly rich informational interaction of conscious agent dynamics to a manageable representation on its interface. Work is ongoing in the study of compression, by means of geometric algebra, from information space (i.e., the state space of the dynamics of CAs) to an efficient interface.

What *is* this world of which we cannot see, with any claim to accuracy, the objective features and structure? The spectacular inadequacy of attempts to solve the hard problem of consciousness indicates that we may have been barking *down* the right tree (i.e., in the wrong direction). At this juncture it seems fair to explore the consequences of the notion that reality may function in the *opposite* direction to physicalism. Indeed, we propose the

Conscious Realism Thesis: The world consists of—and only of—conscious agents.

The hard problem is then reversed: how does the physical world and its laws, as we humans know them, arise in the interaction of conscious agents? In biology we can ask: within the CA formalism, what drives the emergence and evolution of species and their particular interfaces?

Thus we want to discover how to derive physics from consciousness and, in particular, whether quantum theory is a "natural" result of conscious dynamics and what the relation of this dynamics is to classical physics. Here we quote B. Coecke, who suggests that it is the limitations of our classical interfaces that "force" changes of state in a quantum measurement on a (much richer) quantum world: "The above argument suggests that there is some world out there, say the quantum universe, which we can probe by means of classical interfaces. There are many different interfaces through which we can probe the quantum universe, and each of them can only reveal a particular aspect of that quantum universe. Here one can start speculating. For example, one could think that the change of state in a quantum measurement is caused by forcing part of the quantum universe to match the format of the classical interface by means of which we are probing it. In other words, there is a very rich world out there, and we as human agents do not have the capability to sense it in its full glory. We have no choice but to mould the part of that universe in which we are interested into a form that fits the much smaller world of our experiences. This smaller world is what in physics we usually refer to as a space-time manifold" (Coecke 2010). Is this larger world a world of conscious agents?

In certain interpretations, quantum physics has non-spatiotemporal entities. In particular, an extensive study of the question of interpreting these entities by means of a new interpretation of both quantum and relativity theories has been carried out by Aerts and collaborators (see, e.g., Aerts et al. 2018 in this volume, and references therein). They have suggested the "Conceptuality Interpretation" of quantum and relativity theories, which seems to have a similar point of view to that of our conscious realism thesis. In particular, human consciousness is seen just as one form of consciousness – and one which evolved only comparitively recently. By showing how various aspects of quantum theory can be explained by ascribing a conceptual aspect to the interactions of both micro- and macrophysical entities, they have shown how many foundational quantum phenomena and seeming paradoxes may be explained. Such *pan-cognitivism*, as Aerts et. al. have termed it, is in spirit shared by conscious agent theory. For example, in discussing the double-slit

Author's personal copy

experiment, Aerts et. al. suggest that a conceptual measuring device (what they call a "screen mind" in this instance) asks a question of the sort: "What is a good example of an effect produced by an electron interacting with the barrier when it has both slits open?" In the CA formalism, the asking of a question is part of an action of a conscious agent on the world. As Aerts et. al. demonstrate, the response to this question-action is a probability distribution on the device's X, one that models quantum interference effects. It would be fruitful to explore further connections between, on the one hand, the conceptuality interpretation, which assumes the omnipresence in reality of conceptual entities and, on the other hand, conscious agent theory which agrees with that assumption and also gives a precise definition of the structure of consciousness.

In the conceptuality interpretation, observation is in terms of classical entities, which are entirely spatiotemporal. A question arises as to whether it is only such entities which describe the interface of a CA, defined as it is using Markovian kernels on measurable spaces. If the CA's interface consists of classical entities, would the quantum level be the true structure of the world, or would quantum theory describe the relationship of a reduced CA, via its action and perception kernels, to its world? This deep question does not currently have an answer. However, with regard to the classicality of the CA's interface, note that the states of the space X of a CA represent *qualia* and are therefore feeling-conceptual in their very nature.[9] The spatio-temporal aspects of X, if any, might well be a very small aspect of a CA, akin to a coordinate patch locating events that themselves possess a far richer structure. Moreover, a CA may have no spatiotemporal aspect at all (think, e.g., of pure feelings, or of distinguishing between "fruit" and vegetable').[10] The fact that the CA is defined using Markovian kernels on measurable spaces does not preclude the possibilty of quantum behaviour at a deeper level within the dynamics of its interactions with other CAs, so that, e.g., interference and entanglement effects are experienced on, say, a human cognitive interface.[11]

Numerous cognitive science experiments indicate that many aspects of our human conscious experience seem to be closer to quantum-like uncertainties than to neatly distinguishable classical states.[12] Further, from the theoretical perspective of operational logics, it appears that quantum behaviour may well bear a close relation to what is most fundamental in our interaction with reality. See, e.g., D'Ariano et al. (2017), where quantum theory is derived from very basic operational principles. The virtue of the CA formalism is that it provides an avenue to investigate the question of whether quantum theory does indeed model the deepest levels of our conscious reality, or whether there are yet deeper levels to be discovered.

Finally, we can ask: If physics is indeed a result of conscious dynamics, is physical law itself immutable, or is it evolving as consciousness evolves? What, in fact, drives the evolution of consciousness? F. Faggin has suggested (e.g., Faggin 2015) that all the dynamics of this world of consciousness is driven by the desire for *comprehension*, the attainment of new levels of self-knowing, by consciousness itself, of itself.

We look forward to the development of a theory of this process.

[9] It is perhaps an incomplete description to refer to the elements of X as "perceptual" states, as a quale can be a perception, a feeling or even a thought.

[10] When a CA is devoted mostly to spatiotemporality, its functionality would, e.g., be that of a "measuring rod" or a"clock."

[11] If this turns out not to be true, the CA definition may have to be amended to exhibit what we know of quantum behaviour. But the jury is still out on this.

[12] One of the earliest, those of J. Hampton, is referred to in Aerts et al. (2018)

Author's personal copy

Acknowledgements Thanks to the Federico & Elvia Foundation for support for this work. Thanks for fruitful discussions with Federico Faggin, my research colleagues: Donald Hoffman, Chris Fields, Robert Prentner and Manish Singh, and with G. Mauro d'Ariano and Urban Kordes. Thanks also to an anonymous reviewer, whose comments significantly improved the Discussion section. I thank Diederik Aerts, Massimiliano Sassoli de Bianchi and others at VUB, Brussels for hosting the excellent symposium "Worlds of Entanglement" in 2017 and Tomas Veloz and his colleagues for managing the event with aplomb.

References

Aerts, D., Sassoli de Bianchi, M., Sozzo, S., & Veloz, T. (2018). On the conceptuality interpretation of quantum and relativity theories. *Foundations of Science*. https://doi.org/10.1007/s10699-018-9557-z.

Ay, N., & Löhr, (2015). The Umwelt of an embodied agent—A measure-theoretic definition. *Theory in Biosciences*, *134*, 105–116.

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

Chalmers, D. J. (2016). The combination problem for panpsychism. In L. Jaskolla & G. Bruntrup (Eds.), *Panpsychism* (pp. 179–214). Oxford: Oxford University Press.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.

Coecke, B. (2010). Quantum picturalism, *Comtemporary Physics, 51*, 59–83.

D'Ariano, G. M., Chiribella, G., & Perinotti, P. (2017). *Quantum theory from first principles: An informational approach*. Cambridge: Cambridge University Press.

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life*. New York: Simon & Schuster.

Faggin, F. (2015). The nature of reality. *Atti e Memorie dell'Accademia Galileiana di Scienze, Lettere ed Arti*, Volume CXXVII (2014–2015). Padova: Accademia Galileiana di Scienze, Lettere ed Arti. Also, Requirements for a Mathematical Theory of Conscoiusness, *Journal of Consciousness, 18,* 421.

Fields, C. (2018). Private communication.

Fields, C., Hoffman, D. D., Prakash, C., & Singh, M. (2017). Conscious agent networks: Formal analysis and application to cognition. *Cognitive Systems Research*, *47*(2018), 186–213.

Giesler, W. S. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, *27*, 379–402.

Hoffman, D. D. (2000). *How we create what we see*. New York: Norton.

Hoffman, D. D. (2009). The interface theory of perception. In S. Dickinson, M. Tarr, A. Leonardis, & B. Schiele (Eds.), *Object characterization: Computer and human vision perspectives* (pp. 148–165). New York, NY: Cambridge University Press.

Hoffman, D. D. (2019). *The case against reality: Why evolution hid the truth from our eyes*. W. W. Norton, In press.

Hoffman, D. D., & Prakash, C. (2014). Objects of consciousness. *Frontiers of Psychology*, *5*, 577. https://doi.org/10.3389/fpsyg.2014.00577.

Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-015-0890-8.

James, W. (1895). *The principles of psychology*. New York: Henry Holt.

Koenderink, J.J. (2013). World, environment, umwelt, and inner-world: A biological perspective on visual awareness. In *Human vision and electronic imaging XVIII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Huib de Ridder, Proceedings of SPIE-IS&T Electronic Imaging, SPIE* (Vol. 8651, p. 865103) (2013).

Koenderink, J. J. (2015). Esse est percipi & verum factum est. *Psychonomic Bulletin & Review*, *22*, 1530–1534.

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

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Mausfeld, R. (2015). Notions such as "truth" or "correspon- dence to the objective world" play no role in explanatory accounts of perception. *Psychonomic Bulletin & Review*, *22*, 1535–1540.

Mendelovici, A. (2019). Panpsychism's combinaiton problem is a problem for everyone. In W. Seager (Ed.), *The routledge handbook of panpsychism*. London, UK: Routledge.

Novak, M. (2006). *Evolutionary dynamics: Exploring the equations of life*. Cambridge: Belknap Press.

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

Author's personal copy

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

Prakash, C., Stephens, K., Hoffman, D.D., Singh, M. & Fields, C. (2018). Fitness beats truth in the evolution of perception (Under review).

Russel, B. (1959). *The analysis of matter*. Crows Nest: G. Allen & Unwin.

Von Uexküll, J. (2014). Umwelt und Innenwelt der Tiere. In F. Mildenberger & B. Herrmann (Eds.), *Klassische Texte der Wissenschaft*. Berlin: Springer.

Chetan Prakash (Ph.D. Cornell University in Mathematical Physics, 1982) wrote his Ph.D. thesis on Dobrushin uniqueness techniques applied to classical statistical mechanics. Since the 80s he has been working on a rigorous foundation for cognitive science and he co-authored the book "Observer Theory" (with Bruce Bennett and Don Hoffman). Since 2014 he has been working on a new science of consciousness and has co-authored such articles as "Objects of Consciousness" (with Don Hoffman), "Fitness Beats Truth in the Evolution of Perception" (with Hoffman, Stephens, Singh and Fields) and "Conscious Agent Networks: Formal Analysis and Application to Cognition" (with Fields, Hoffman and Singh). This research uses rigorous mathematical analyses to demonstrate the falsity of the commonly held belief that we have evolved to perceive, with ever-increasing accuracy, a fixed, "objective" reality. This supports the idea that it is consciousness that gives rise to our "reality," as against the opposite belief, prevalent in today's scientific community, that material processes (such as brains) produce consciousness. Chetan's current research is on this "reverse hard problem of consciousness". Dr. Prakash is Professor Emeritus at California State University, San Bernardino and currently functions as an independent researcher.