The Interface Theory of Perception

DONALD D. HOFFMAN

INTRODUCTION

Our biological organs — such as our hearts, livers, and bones — are products of evolution. So too are our perceptual capacities — our ability to see an apple, smell an orange, touch a grapefruit, taste a carrot and hear it crunch when we take a bite. Perceptual scientists take this for granted. But it turns out to be a nontrivial fact with surprising consequences for our understanding of perception. The evolution of perception may be taken for granted, but what this evolution entails is not yet well understood.

The evolution of perception profoundly affects the answers to questions that are fundamental to the science of perception: What is the relationship between one's perceptions and objective reality, i.e., reality as it is when one does not observe? Do our perceptions accurately estimate objective reality? Why do our actions transform our perceptions in systematic and predictable ways? What do such transformations entail about the structure of objective reality? What are psychophysical laws? Why do they have the form that they do? What are illusions? What are hallucinations? How precisely do they differ from normal perceptions? What is the content of a perceptual experience?

The evolution of perception also profoundly affects the answers to questions that are fundamental to cognitive neuroscience more generally: What is the relationship between mental states and neural activity? Do neural states and processes have causal powers? Do they cause conscious experiences and other mental states?

It is widely assumed by vision scientists that evolution shapes our perceptions to accurately estimate true properties of reality.

For instance, Palmer (1999) says, “Evolutionarily speaking, visual perception is useful only if it is reasonably accurate ... Indeed, vision is useful precisely because it is so accurate. By and large, what you see is what you get. When this is true, we have what is called veridical perception ... perception that is consistent with the actual state of affairs in the environment. This is almost always the case with vision...” (emphasis his).

Knill et al (1996, p. 6) say, “Visual perception ... involves the evolution of an organism’s visual system to match the structure of the world and the coding scheme provided by nature.”

Marr (1982, p. 340) says, “We ... very definitely do compute explicit properties of the real visible surfaces out there, and one interesting aspect of the evolution of visual systems is the gradual movement toward the difficult task of representing progressively more objective aspects of the visual world.”

The intuition is that those of our ancestors who saw more accurately enjoyed a competitive advantage over those who saw less accurately, and were therefore more likely to pass on their genes that coded for more accurate perceptions. We are the result of thousands of generations of this process, and thus we can be confident that, in the normal case, our perceptions accurately estimate those properties of reality that are critical for our survival and reproduction.

Geisler and Diehl (2003) say this succinctly: “In general, (perceptual) estimates that are nearer the truth have greater utility than those that are wide of the mark.”

Trivers (2011) spells it out a bit more: “...our sense organs have evolved to give us a marvelously detailed and accurate view of the outside world—we see the world in color and 3-D, in motion, texture, nonrandomness, embedded patterns, and a great variety of other features. Likewise for hearing and smell. Together our sensory systems are organized to give us a detailed and accurate view of reality, exactly as we would expect if truth about the outside world helps us to navigate it more effectively.”

This intuition is compelling but, as we shall see, false.

Monte Carlo simulations of evolutionary games demonstrate that perceptions which accurately estimate reality never outcompete perceptions of equal complexity which do not estimate reality but are instead tuned to the relevant fitness functions (Mark et al., 2010; Hoffman et al., 2013; Marion, 2013; Mark, 2013).

The key idea here is the *fitness function*. What is the fitness conveyed by, say, a piece of raw beef? The answer depends on the organism, its state, and its action. For a hungry cheetah looking to eat, the beef enhances fitness. For a sated cheetah looking to mate, it does not. And for a cow looking to do anything, it does not. Thus a fitness function depends not just on the state of objective reality, but also, and crucially, on the organism, its state and action. Fitness functions, not objective reality, are the coin of the realm in evolutionary competition.

The results of Monte Carlo simulations are now buttressed by the *Fitness-Beats-Truth (FBT)* Theorem: For an infinitely large class of generically chosen worlds, for generically chosen probabilities of states on the worlds, and for generically chosen fitness functions, an organism that accurately estimates reality is *never,* in an infinite class of evolutionary games, more fit than an organism of equal complexity that does not estimate objective reality but is instead tuned to the relevant fitness functions.

The FBT Theorem says the probability is low, approaching zero, that any of our perceptions estimate true properties of objective reality. More deeply, it says the very *predicates* of our perceptions — predicates such as space, time, physical objects, position, momentum, and 3D shape — are the *wrong language* to describe reality. The problem is not that our perceptual estimates are a tad off here or there and need minor corrections. The problem is that no correction is possible because the language of physical objects in spacetime cannot possibly describe reality as it is.

This point is fundamental. Current models of perception — such as Bayesian, inverse optics, ecological optics, and enactive models — disagree on much, but they all agree that perceptual predicates such as space, time and shape are appropriate to describe objective reality. The FBT Theorem says that they are wrong.

But how could perception be useful if it does not, and could not, describe objective reality? How could failing to see objective reality grant a competitive advantage? The interface theory of perception (ITP) answers this question; its answer entails radical, and empirically testable, answers to the panoply of questions that opened this section (Fields 2014; Hoffman 1998; 2009; 2011; 2012; 2013; Hoffman and Prakash 2014; Hoffman and Singh 2012; Hoffman et al. 2013; Hoffman et al. 2015a, 2015b; Koenderink 2011; 2013; Mark et al. 2009; Mausfeld 2002; Singh and Hoffman 2013; see also von Uexküll (1909; 1926; 1934) for his related idea of an Umwelt).

ITP: AN INFORMAL INTRODUCTION

This section describes ITP informally, the next mathematically.

ITP says that our perceptions are not a window onto objective reality, but instead they are more like the windows interface of a computer. Suppose you are editing a text file, and the icon for that file is green, rectangular and in the center of the desktop. Does this mean that the text file itself is green, rectangular and in the center of the computer? Of course not. Anyone who thinks so completely misunderstands the purpose of the interface. The shapes, colors and positions of its icons are not meant to depict the real shapes, colors and positions of files in the computer. Indeed, these are the wrong predicates; files have no colors or well-defined shapes.

Instead the purpose of the interface and its icons is to *hide* the real nature and complexity of the computer, and to provide simple tools that allow the user to edit files and photos without the burden of dealing directly with transistors, voltages, magnetic fields and megabytes of software. Good luck trying to craft an email by directly manipulating voltages in a computer. Fortunately, the interface lets you manipulate them without knowing anything about them, so you can easily write and send that email.

According to ITP, space-time as we perceive it is our species-specific desktop, and physical objects as we perceive them are species-specific icons in that desktop. Our perceptions of space-time and physical objects

are not an insight into objective reality. Instead, they are a species-specific interface that hides objective reality and guides adaptive behaviors. Perception is not about seeing truth, it's about having kids.

The stakes are high. If ITP is right, then space-time is doomed. Not only are our perceptions in terms of space and time not an insight into objective reality, but even more importantly the very predicates of space and time are almost surely the *wrong predicates* to describe objective reality. Thus ITP makes a prediction that goes to the heart not just of theories of perception, but also of theoretical physics. ITP predicts that physicists will find that space-time is doomed, that the language of space-time is the wrong language in which to formulate the deepest theories of physics. If this prediction is wrong, then ITP is almost surely disconfirmed. Some physicists indeed claim that space-time is doomed and must be replaced, although they don't yet know what should replace it (e.g., Giddings 2015; Arkani-Hamed 2015, at https://www.youtube.com/watch?v=KyRO8Wv4BaY).

If ITP is right, then physical objects are doomed. Not only are our perceptions in terms of physical objects not an insight into objective reality, but even more importantly the very predicates of position, momentum, spin and shape are almost surely the wrong predicates to describe objective reality. Thus ITP makes a clean prediction: No physical object has a definite value of any dynamical physical property (such as position, momentum, spin) when it is not observed. If any experiment demonstrates otherwise, then ITP is disconfirmed. One might argue that this prediction by ITP is not really a prediction at all. How could one possibly do an experiment to show that a physical object has no position *when it is not observed*? This seems as pointless as asking how many angels can dance on the head of a pin.

But one *can* do such experiments, they have been done repeatedly, and the result every time is as predicted by ITP: No physical object has a definite value of any dynamical physical property when it is not observed. Among these remarkable experiments are, e.g., Ansmann et al., (2009), Cabello et al., (1996), Fuchs (2010), Giustina et al., (2013), Pan et al., (2000), Rowe et al., (2001), Salart et al. (2008), Weihs et al., (1998). A helpful introduction to these experiments is Mermin (1985).

If ITP is right, then *causality* of physical objects is doomed. When a bat hits a ball, we naturally assume that the subsequent motion of the ball is *caused* by its collision with the bat. ITP entails that this assumption is false. Bat and ball are simply perceptual icons employed by *H. sapiens* to guide adaptive behavior. They are not insights into the causal structure of objective reality.

To understand this claim, it's helpful to return to the interface metaphor. If one drags a file icon to the trash icon, it certainly *seems* like the interaction of the two icons *causes* the file to be deleted. This appearance is of course an illusion. There is no feedback to the computer from the apparent movement and interaction of icons on screen. For most purposes it is a harmless, even useful, fiction to attribute causal powers to the icons. But for someone who wants to understand the workings of the computer, this fiction is no longer harmless.

Similarly, for most everyday purposes, indeed for most scientific purposes, it is a harmless, even useful, fiction to attribute causal powers to physical objects in space-time. For instance, for most research in neuroscience it is a harmless fiction to assume that neurons have causal powers, and that neural activity causes our behaviors and conscious experiences. But for someone who wants to understand the hard problem of consciousness — namely, how our conscious experiences are related to brain activity — this fiction is no longer harmless, and has blocked progress for centuries. ITP makes a clear prediction: Neural activity causes none of our behaviors and none of our conscious experiences. If experiments prove otherwise, then ITP is disconfirmed.

A few objections and replies help to explore key ideas of ITP.

Objection 1. "If you think that the rattle snake over there is just an icon of your interface, and has no causal powers, then why don't you walk over and grab it? After you're dead, and your theory with you, we'll know for sure that the snake is not just an icon, and that its bite indeed has causal powers."

Reply 1. The answer is that I would not grab the snake for the same reason I would not carelessly drag my file icon to the trashcan. Not because I take the icon literally — the file is not green and rectangular — but I do take it seriously. If I'm not careful, I could lose months of work.

So this objection rests on the following logical mistake:

Premise: I must take my perceptions seriously.
Conclusion: I must take my perceptions literally.

The mistake is clear for icons on the desktop. We must take them seriously (or we could carelessly lose our work), but must not take them literally (files are not literally green and rectangular). But it is the same mistake for snakes, cliffs and other physical objects. Our perceptions in terms of snakes and cliffs have been shaped by evolution to keep us alive long enough to reproduce, so we must take them seriously. Those who don't take their perceptions seriously also have a penchant for dying young. But logic does not require that we take them literally, and the theory of evolution entails that we should not take them literally. Thus the idea that more accurate perceptions are fitter perceptions has its genesis in an error of logic and a mistaken understanding of evolution.

Objection 2. "That snake over there is not just an icon, because everyone who looks over there sees the snake. So the snake must be part of objective reality, and is not just my perceptual icon."

Reply 2. This objection rests on the following logical mistake:

Premise: We all agree that we see a snake.
Conclusion: Therefore the snake is not just a perceptual icon.

The conclusion does not follow. For instance, one reason we might all agree that we see a snake is that we are all members of the same species and, in consequence, our perceptual systems produce similar icons in similar contexts. Consensus is just consensus. It does not logically entail the objective reality of what we agree we perceive.

Moreover there are clear counterexamples to this logic. We all agree that we see a 3D cube when we look at a line drawing of the famous Necker cube. But the drawing is flat, so the 3D cube that we each see is just a

construct of the perceptual system — a perceptual icon. We all agree that we see a 3D cube, but the reason we agree is that we each construct a similar perceptual icon. I see the cube that I construct, and you see the cube that you construct. The same holds true of the snake. I see the snake that I construct, and you see the snake that you construct. There is something that exists whether or not you or I look. But that something is not a snake and in no way resembles a snake. A snake is just the symbol that you and I — as humble members of a particular species with its own inherent perceptual limitations — have been shaped by evolution to construct.

Objection 3. "If natural selection did not design our senses and brain to construct a relatively accurate model of reality, then how is it we can land a spacecraft on Mars with pinpoint accuracy, or put satellites in space so finely tuned with relativity theory that they can detect a device on earth within less than a meter of accuracy? That would seem to be fairly strong support for the claim that our senses, coupled to the methods of science, are coming ever closer to an accurate model of reality—the way it actually is. If our science was way off then how is it we can do such things?"

Reply 3. This is a central question. It is answered by the *Invention of Symmetry Theorem* (*IOS Theorem*, Hoffman et al., 2015a), which states that one can have perceptions of 3D space and 1-D time together with perfectly coordinated actions in that perceived space and time that are entirely predictable to arbitrary accuracy, and yet this entails absolutely nothing about the nature of objective reality, except to put a lower bound on the cardinality of its set of states. The IOS Theorem is not restricted to 3D space and 1-D time. The IOS Theorem holds for arbitrary groups of any dimension.

So in fact there is a precise theorem that answers this objection. The IOS Theorem entails that natural selection can shape our perceptions to be tuned to fitness and not to reality as it is — i.e., so that our perceptions are akin to a user interface — and yet we can have very predictable (and fitness enhancing) perception-action interactions via that interface, interactions such as landing a spacecraft on Mars.

Together the IOS Theorem and FBT Theorem flatly contradict the claim that the regularities of our perceptions — regularities such as the Euclidean group of three-dimensional space — are evolutionarily internalized versions of regularities in the objective world (e.g., Shepard 1987, 1994, 2001; Tenenbaum and

Griffiths 2001; Chater and Vitanyi 2003). Instead, the regularities of our perceptions are an evolutionarily designed interface that guides adaptive behavior and hides the true regularities of the objective world.

The IOS Theorem applies to shape perception. Pizlo (2012; 2015) has argued that symmetry, together with planarity and compactness, allows our perceptual systems, in a principled fashion, to accurately reconstruct the objective 3D shapes of objects. As he puts it, "To summarize, in the new approach advocated here, 3D shape reconstruction relies on three main shape constraints: symmetry, planarity, and compactness. These three constraints become the essential element of a "new simplicity principle," according to which the perceived shape corresponds to the minimum of a cost function that represents the interaction between a 2D shape on the retina (produced by a 3D shape "out there") and a priori constraints applied to the 3D shape percept. Note that this new simplicity principle is used because it leads to accurate 3D reconstructions, not because objects "out there" are simple." (Pizlo 2012, section 5.1). The IOS Theorem entails that no symmetry enjoyed by any of our perceptions, including our perceptions of shape, constrains the symmetries, if any, that in fact obtain in objective reality. Symmetry cannot be used to recover veridical shapes. Any perceived symmetry can be accommodated by an objective world which in fact fails to have that symmetry, just so long as the cardinality of the possible states of that objective world is sufficiently large. In short, it is a theorem that one cannot in general accurately reconstruct any elements of objective reality using symmetry, planarity, and compactness.

Objection 4. "You claim to use the theory of evolution to prove that our perceptions of space-time and physical objects do not reflect objective reality. But the theory of biological evolution *assumes* that there really are physical objects in space-time, objects such as organisms, DNA, food resources, and so on. So you are using evolution to disprove evolution. You are caught in a logical blunder, and have refuted yourself."

Reply 4. Not at all. At the heart of evolutionary theory is an algorithmic core — variation, selection, and retention — that has been called "universal Darwinism" (Dawkins 1983; Dennett 1995; von Sydow 2012). It is this algorithmic core that is captured in the formalism of evolutionary game theory, evolutionary graph theory, and genetic algorithms. And it is this algorithmic core that Dennett has argued is the "universal

acid" of evolutionary theory that extends to domains beyond biological evolution and can fundamentally transform them. For instance, universal Darwinism has been applied to the spread of "memes" in human culture.

It is this algorithmic core that is used by the FBT Theorem to conclude that our perceptions of space-time and physical objects do not reflect objective reality. Thus the acid of universal Darwinism can be applied to the theory of biological evolution itself. The consequence is that it etches away superfluous assumptions of the theory, such as the assumption of the objective reality of physical objects and space-time. These assumptions were part of the scaffolding that helped to formulate the theory of evolution. But now that the theory is formed, its logical core can be used to discover what is essential to the theory and what can be discarded. This is part of the power of the scientific method. Our theories are ladders to new levels of understanding, and sometimes a new level of understanding leads us to kick away the very ladder that led to it.

To put this another way, suppose we want to ask the question, "Do our perceptions in terms of space-time and physical objects reflect reality as it is?" And suppose we ask if the theory of evolution can help us answer this question. If it turned out that evolutionary theory could not help us, because the only answer it could possibly give is "Yes," then surely we would want to turn elsewhere to find a more powerful framework with the chance of giving us a genuine answer. But, to its credit, the theory of evolution is indeed powerful enough to give us a genuine answer, and that answer, remarkably, requires us to reconsider how we think about that theory in the first place. For instance, it entails that DNA does not exist when it is not perceived. Something exists when we don't look that causes us, when we do look, to perceive DNA, but, whatever that something is, it's not DNA. Indeed it's not in spacetime. DNA is just the representation that we, as humble members of the species *H. sapiens*, have been shaped by evolution to construct. We must not mistakenly take the limits of our evolutionarily endowed perceptions to be insights into the nature of objective reality.

Objection 5. "The question of whether our perceptions are truthful is irrelevant to scientific theories of perception. Those theories aim to understand the internal principles of a biological system" (see, e.g., Mausfeld 2015).

Reply 5. There are two large classes of questions one can ask about biological perception. *Proximate* questions inquire into how biological systems currently operate — e.g., how we compute depth from stereo and motion, how we achieve approximate color constancy in a wide variety of ambient illuminations. *Ultimate* questions inquire into why biological systems operate as they do — why did we evolve to see depth in three dimensions, and why did we evolve to have the color perceptions that we do? Both kinds of questions are essential in the science of perception.

Proximate questions require proximate explanations, describing the mechanisms, algorithms and internal principles of the *current* operation of the perceptual system. Ultimate questions require ultimate explanations, describing the *evolutionary pressures* that forged these mechanisms, algorithms, and internal principles.

The question of whether our perceptions are truthful, i.e., accurately describe objective reality, is an ultimate question that requires an evolutionary explanation. Discovering that the answer is "No" fundamentally changes how we think about the proximate questions. Far from being irrelevant, this answer fundamentally transforms all aspects of perceptual theory.

Objection 6. "You claim that none of our perceptions reflect objective reality. That means that all of our perceptions are illusions, which is a reductio of your theory."

Reply 6. Illusions are a critical source of insight in modern perceptual theory, because they sometimes reveal assumptions built into the computations that underly normal perception. Palmer (1999, p 313) describes the standard account of illusions as follows: *"*... veridical perception of the environment often requires heuristic processes based on assumptions that are usually, but not always, true. When they are true, all is well, and we see more or less what is actually there. When these assumptions are false, however, we perceive a situation that differs systematically from reality: that is, an illusion*"*.

This standard account assumes that in the normal case our perceptions depict reality accurately, i.e., veridically, and that they achieve this accuracy via computations that are based on assumptions about

reality, assumptions that are usually true. On rare occasions the assumptions don't match reality and, in consequence, the resulting perceptions don't match reality, leading to an illusion. So, on the standard account, an illusion is a non-veridical perception that most normal perceivers have in certain unusual contexts.

This standard theory of illusions clearly cannot be endorsed by ITP, because ITP says that *none* of our perceptions are veridical, and this would entail, on the standard theory of illusions, that *all* of our perceptions are illusory. It would be unhelpful for ITP to say that all perceptions are illusory. There is, after all, some kind of important distinction between perceptions that we deem normal and those we deem illusory. ITP owes us a new account of this important distinction.

Here it is — the ITP theory of illusions: An illusion is an *unadaptive* perception that most normal perceivers have in certain unusual contexts.

For instance, when one sees a standard depth illusion, such as the Necker cube display, ITP says that it is an illusion not because the perception of depth fails to match objective reality (it *never* matches, according to ITP), but because the perception of depth and shape incorrectly indicates that certain interactions with the environment are possible (e.g., grasping a 3D cube). Attempts to perform these actions would fail, and thus the perception fails to guide adaptive behavior. This explanation of illusions follows naturally from the basic difference between ITP and standard theories of perception: Where standard theories claim that our perceptions are, in the normal case, *accurate representations of reality*, ITP says that they are *adaptive guides to behavior*. So the standard theory naturally says illusions are inaccurate representations of reality, whereas ITP says they are unadaptive guides to behavior.

There are, of course, many more objections to ITP that must be addressed. But the ones just discussed help fix the basic ideas of ITP.

ITP: A FORMAL INTRODUCTION

Evolutionary game theory, evolutionary graph theory, and genetic algorithms are powerful tools to evaluate the relative fitness of competing strategies (Hofbauer and Sigmund 1998; Mitchell 1998; Lieberman et al. 2005; Nowak 2006; Samuelson 1997; Sandholm 2007). To understand ITP rigorously, and to precisely evaluate the fitness of ITP vis-à-vis veridical and other types of perceptions, it is helpful to formalize ITP as a *perceptual strategy* and to place it in the context of all possible perceptual strategies. This is the goal of this section.

Intuitively, a perceptual strategy is a map from objective reality to perceptual experiences. Already we are in trouble, because the notions of *objective reality* and *perceptual experiences,* and their relationship, have provoked debate for centuries (e.g., Brewer 2011; Byrne and Hilbert 2003; Campbell and Cassam 2014; Coates 2007; Fish 2009, 2010; Searle 2015).

Here the goal is not to take sides in this debate. To the contrary, we want to devise a precise classification of *all possible* perceptual strategies. This means a classification that covers all possible notions of objective reality, perceptual experiences, and their relationship. This task might seem impossible. But mathematics has the right tools. It allows us to describe these notions and their relationship abstractly, without a priori commitment to any particular account.

In what follows I will present an abstract description and the classification it provides. The virtue of this description and classification is not that it is necessarily correct and complete — although correctness and completeness are of course the goal — but that it is precise. Anyone who wishes to challenge it has a precise target. A serious challenge would state precisely what is wrong with the mathematical description or classification, and propose a mathematically precise alternative.

If the classification succeeds, then *every* philosophical and scientific theory of the relationship between objective reality and perceptual experiences should fall somewhere within it. In this case, the classification provides a much-needed framework for comparing different theories. However, if someone produces a

theory of perception and reality whose structure lies outside the classification, this will require modification of the classification or of its underlying mathematical description.

We begin by recognizing that an organism, such as a person, is part (or, according to metaphysical solipsists, all) of *total* reality, whatever total reality might be. Let us represent the *organism* by a set *O* and *total reality* by a set W^T. Let *objective reality* be represented by the set $W = W^T - O$; objective reality is total reality excluding the organism. (Metaphysical solipsists would say that *W* is the empty set; we will assume it is not.) Let us represent the *perceptual experiences* of the organism by a set *X*. Each point of *X* represents a specific perceptual experience of the organism.

In the most general case, we assume no a priori structure on the sets W^T, *W*, *O*, and *X* — no metrics, topologies, orders or algebras. A specific theory might of course posit such structures and thereby place itself in some restricted part of the classification that follows.

In the most general case, we assume that the relationship between *W* and *X* is some map, $P : W \to X$. This places no restrictions on the nature of *P*, such as continuity, measurability, or preservation of any structures. Moreover it allows, e.g., the special case that $X \subset W$ and that *P* has the form $P : W \times X \to X$.

We can now state

Definition 1: A *perceptual strategy* is a mapping $P : W \to X$, where *W* is a set representing objective reality and *X* is a set representing the perceptual experiences of an organism.

I now classify the kinds of perceptual strategies, beginning with the most general.

The scientific study of perceptual evolution requires, at least, that *P* systematically relates perceptual events, such as tasting chocolate or smelling smog, to events in objective reality, whatever they might be.

Otherwise perceptual outcomes, being unmoored to objective reality, can neither inform the behavior of organisms nor constrain the theories of scientists.

Mathematically, this means assuming that perceptual events have certain properties. First, if there is a perceptual event *tastes like chocolate* then there is also a perceptual event *doesn't taste like chocolate*. So every event entails the existence of another event that is its complement. Second, if there is a perceptual event *tastes like chocolate* and a perceptual event *feels like water*, then there is a perceptual event *tastes like chocolate and feels like water*, and a perceptual event *tastes like chocolate or feels like water.* So every pair of events entails the existence of a corresponding conjunction event and of a corresponding disjunction event.

Mathematically, this cashes out as follows. A perceptual event is a subset E of X. The collection of all such events is denoted $\mathcal{X}$ ("curly X"), and is closed under complement (if E is in $\mathcal{X}$ then the complement of E is in $\mathcal{X}$), union (if E is in $\mathcal{X}$ and F is in $\mathcal{X}$ then the union of E and F is in $\mathcal{X}$) and, by de Morgan's laws, also intersection (if E is in $\mathcal{X}$ and F is in $\mathcal{X}$ then the intersection of E and F is in $\mathcal{X}$). If one allows a countable collection of events and closure under countable union, then the resulting structure of events is called a σ-algebra. This structure provides a framework of events on which one can define probability measures.

The requirement that P systematically relates perceptual events to events in objective reality then means that P respects the σ-algebra of events $\mathcal{W}$ on W and $\mathcal{X}$ on X. If P is a function, then this means that every event E in $\mathcal{X}$ is pulled back by P to an event $P^{-1}(E)$ which is in $\mathcal{W}$. Such a function is called a *measurable function*. This models the case in which there is no dispersion. To model dispersion we can generalize from functions to kernels. In what follows, for simplicity I will focus on functions; the pattern of results is the same for kernels.

Our most general perceptual strategy, then, is one that puts no constraint on P other than measurability. Thus we define

Definition 2. An *interface* perceptual strategy is a measurable mapping $P : W \to X$.

Note that this definition puts no other constraint on the relationship between *W* and *X*. In particular, it does not require that *X* is a subset of *W*, nor does it require that *P* is a homomorphism (i.e., that *P* respects) any structures on *W* other than its σ -algebra, e.g., orders, topologies, or metrics. If *P* does not respect these other structures, this limits what one can easily infer about *W* from *X*. For instance, if *X* has a 3D space with a Euclidean metric, then one cannot infer from this alone that *W* has a space of some dimension with a Euclidean or other metric. Thus interface strategies need not be veridical. Other than the event structure, any structures on perceptual experiences, such as metrics and orders, need not indicate anything at all about the structure of the objective world.

This is counterintuitive to most perceptual scientists, who assume that perception is at least veridical in the sense that the structures of our perceptions are systematically related to structures in the objective world. This intuition is captured by the following

Definition 3. A *critical realist* perceptual strategy is an interface strategy that is also a homomorphism of all structures on *W*.

The critical realist strategies are a proper subset of the interface strategies. We will call an interface strategy that is not a critical realist strategy a *strict interface* strategy. A strict interface strategy is non veridical.

Even critical realist strategies are not veridical enough for some perceptual scientists, who maintain that at least some of our perceptual experiences are in fact part of objective reality, not merely homomorphic to objective reality. For instance, Pizlo (2015) asserts that our perceptual systems recover the true 3D shapes of objects in the world, so that our perception of these shapes is identical to the shapes themselves. He extends this idea to other properties of objects as well: "… the 3D symmetrical shapes of objects allow us not only to perceive the shapes themselves, veridically, but also to perceive the sizes, positions, orientations, and distances among the objects veridically."

To capture this intuition, we first define a strategy in which all of our perceptual experiences are in fact part of objective reality

Definition 4. A *naive realist* perceptual strategy is a critical realist strategy for which X is a subset of W.

The naive realist strategies are a proper subset of the critical realist strategies. We will call a critical realist strategy that is not a naive realist strategy a *strict critical realist* strategy. A strict critical realist strategy has no perceptual experiences that are in fact part of objective reality, but the relations among perceptions of a critical realist are nevertheless homomorphic to the relations on W.

The naive realist strategy has *all* perceptual experiences being part of objective reality. This seems too strong to some perceptual researchers who argue that even if some perceptual experiences such as 3D shape are part of objective reality, nevertheless other perceptual experiences such as color are not. Objects in reality have shapes but not, strictly speaking, colors. Instead they might have related properties such as reflectances or spectral distributions. To capture this intuition we define

Definition 5. A *hybrid realist* strategy is a critical realist strategy for which a subset X' of X is a subset of W.

Finally, for completeness we consider the most restrictive strategy in which the perceiver sees all and only objective reality.

Definition 6. An *omniscient realist* strategy is a naive realist strategy in which $X = W$.

We know of no perceptual scientists who are omniscient realists. But we include this strategy to make our collection of strategies comprehensive. We don't want to rule out a strategy a priori before evaluating its evolutionary potential.

Given this nested collection of perceptual strategies, we can now define ITP formally: ITP asserts that our perceptual strategy is a strict interface strategy.

The most common theories of perception today assert that the perceptual strategy of *H. sapiens* is a hybrid realist strategy.

So what does evolution assert?

The Fitness-Beats-Truth Theorem

We can use evolutionary game theory to find out which perceptual strategies are favored by evolution. We can, for instance, create randomly generated worlds in which there are territories that have random quantities of various resources. We can choose various kinds of payoff functions that relate the quantities of each resource to the fitness payoffs an organism receives if it acquires or consumes that resource. Then we can place artificial organisms in these worlds, each organism having a perceptual strategy, and let them compete for resources.

To be concrete, in one game we might have a world with three territories, each territory having two resources, which we could think of as food and water. The quantity of food can vary from 0 to 100, as can the quantity of water. The probability distribution of food and water can be varied randomly. The fitness payoffs might grow linearly with the amount of resources, or it might vary nonlinearly according to a bell curve. We can place an artificial organism using a strict interface strategy in competition with another artificial organism using a veridical strategy, say an omniscient realist strategy. We can allow them to compete hundreds of thousands of times, with the food and resource quantities randomly chosen each time.

We can then vary the number of territories and the number of resources per territory to study the effects of complexity on the evolutionary outcomes.

Evolutionary games model frequency-dependent selection: the fitness of a strategy is not time-invariant, but varies with the proportion of individuals in the population that use each strategy (Allen and Clarke 1984; Hofbauer and Sigmund 1998; Nowak 2006; Samuelson 1997; Sandholm 2007).

For instance, consider the strategies of hunter gatherers who share their daily catch. Some work hard to hunt and gather, whereas others free-load and simply eat what others provide (Barnard and Sibly 1981). If most work hard, then free-loaders do well; but as the proportion of free-loaders increases, the fitness of their strategy declines until, in the limit where everyone free-loads, everyone starves.

Another example is Batesian mimicry, in which a harmless species avoids being eaten by resembling a dangerous species. In locales frequented by the dangerous species, even poor mimics avoid predation; in locales less frequented by the dangerous species, only good mimics escape being dinner (Harper and Pfennig 2007).

Evolutionary game theory assumes infinite populations of organisms, each having a fixed strategy. It also assumes complete mixing: Organisms are chosen at random to compete in games, so that any pair (or triple — in three-way competitions, etc.) of organisms is equally likely to be picked for competition. Each organism receives a payoff from each of its competitions. This payoff is equated with fitness, i.e., with reproductive success. The result is natural selection: strategies that compete better in games reproduce more quickly and thus outcompete other strategies.

In evolutionary game theory, natural selection is modeled formally by a differential equation called the *replicator equation* (Bomze 1983; Taylor and Jonker 1978). Suppose that in each competition n strategies interact. Let a_{ij} be the payoff to strategy i when it competes with strategy j. Let $[a_{ij}]$ denote the $n \times n$ matrix of such payoffs for all possible competitions between strategies. And let x_i denote the frequency of strategy i in the population of organisms. The expected payoff for strategy i is then $f_i = \sum_{j=1}^{n} x_j a_{ij}$. The average payoff for all strategies is $\varphi = \sum_{i=1}^{n} x_i f_i$. By equating payoffs with fitness, we obtain the replicator equation: $\dot{x_i} = x_i(f_i - \varphi)$. Here $i = 1,\ldots,n$ and $\dot{x_i}$ denotes the time derivative of the frequency of strategy i.

If there are just two strategies, then strategy 1 dominates if $a_{11} > a_{21}$ and $a_{12} > a_{22}$. Strategy 2 dominates if these inequalities are reversed. Strategies 1 and 2 are bistable if $a_{11} > a_{21}$ and $a_{12} < a_{22}$. Strategies 1 and 2 coexist if $a_{11} < a_{21}$ and $a_{12} > a_{22}$. Strategies 1 and 2 are neutral if $a_{11} = a_{21}$ and $a_{12} = a_{22}$ (Nowak 2006).

Monte Carlo simulations of such evolutionary games give a clear verdict: Veridical perceptions go extinct when they compete against strict interface perceptions of equal complexity, i.e., the same number of perceptual states in X (Mark, Marion, & Hoffman 2010; Marion 2013; Mark 2013). In many cases the strict interface strategy wins even when its number of perceptual states is substantially less than that of the veridical strategy.

One might assume that strict interface strategies could win simply because they might take less time and resources. After all, if you don't have to compute all the details of reality, then that can be a valuable savings. But the real reason for the advantage of strict interface strategies is much deeper than this: For evolutionary competitions, truth is *irrelevant*. Only the fitness function is relevant to the outcome of the competition. Resources spent estimating the truth are, generically, resources not spent estimating the only thing that matters: fitness payoffs. If the fitness payoffs vary non-monotonically with structures in the objective worlds, as they certainly do generically and even more so when organisms must maintain homeostasis, then fitness and truth decouple completely and an organism is wasting its time if it estimates truth. The simulations show that strict interface strategies dominate not simply because they can be cheaper, but because they are tuned to fitness and waste no resources estimating anything else.

However, simulations are simulations, not theorems. One could argue that the random sampling still might have missed important cases where estimating truth gives evolutionary advantage.

This issue has been settled — for an infinite class of worlds, resource distributions and fitness functions — by a theorem proven by Chetan Prakash (Prakash, Hoffman & Singh 2016).

Fitness-Beats-Truth (FBT) Theorem: For any world whose states form a compact regular Borel space, for any given perceptual map, for all possible fitness functions, and for all *a priori* probabilities of world states that are absolutely continuous with respect to the Borel measure, the probability that a strict interface strategy strictly dominates an omniscient realist strategy whose perceptual space X is of equal size $|X|$, is at least $(|X|-3)/(|X|-1)$. As this size increases, this probability becomes arbitrarily close to 1: in the limit, a strict interface strategy will generically strictly dominate an omniscient realist strategy, so driving the latter to extinction.

The set of worlds covered by this theorem includes any compact subset of any Euclidean or nonEuclidean space.

Many researchers assume that the more complex the world and the organism's perceptions, the more helpful it will be to have veridical perceptions of that world. The FBT Theorem shows that this assumption is backward. As the organism gets more complex, the probability that veridical perceptions will escape extinction goes to zero. Moreover, more complex worlds do not make veridical perceptions more fit.

The FBT Theorem is, for many, counterintuitive. For instance, Fodor (2000) says, "There is nothing in the ''evolutionary'', or the ''biological'', or the ''scientific'' worldview that shows, or even suggests, that the proper function of cognition is other than the fixation of true beliefs" (p. 68). Fodor's claim is widely assumed to be true by vision scientists and cognitive neuroscientists more generally.

However, Pinker (2005) points out, "Members of our species commonly believe, among other things, that objects are naturally at rest unless pushed, that a severed tetherball will fly off in a spiral trajectory, that a bright young activist is more likely to be a feminist bankteller than a bankteller, that they themselves are above average in every desirable trait, that they saw the Kennedy assassination on live television, that fortune and misfortune are caused by the intentions of bribable gods and spirits, and that powdered rhinoceros horn is an effective treatment for erectile dysfunction. The idea that our minds are designed for truth does not sit well with such facts." Pinker goes on to give other good reasons for believing that the function of cognition is something other than believing true things. For instance, the best liar is the one who

believes his own lies. The FBT Theorem clearly sides with Pinker: Because fitness and truth generically diverge, natural selection generically shapes us away from truth in order to make us more fit.

Perceptual Agents

ITP asserts, as we have said, that natural selection favors strict interface strategies, and that it is thus overwhelmingly likely that none of our perceptions are veridical; indeed the very predicates employed by our perceptual systems—space, time, objects, shapes, colors, motions, positions—are almost surely the wrong predicates to describe objective reality.

But ITP must say more. Perception is an active process. We don't just sit around passively receiving unsolicited perceptions from the objective world. We decide how we will act on the world, and we develop expectations about the perceptual consequences of our actions. ITP must provide a formalism that can deal with all this.

It does. The formalism is intended to be a simple, yet universal, language for describing all aspects of perception, decision and action. It follows the example set by Alan Turing in defining a Turing machine. The Turing machine is an extremely simple formalism—a finite set of symbols, a finite set of states, a start state, a set of halt states, and a finite set of transition rules. Yet, the Church-Turing Thesis claims that every effective procedure can be instantiated by some Turing machine. No counterexample has ever been successfully offered against the Church-Turing Thesis, and the Turing machine is widely accepted as a universal formalism for computation. It is a key pillar of theoretical computer science.

ITP proposes the formalism of *Perceptual Agent Theory* (*PAT)*. One component of this formalism we have seen already: a perceptual map $P : W \to X$ that is a measurable function. However, PAT allows for dispersion, and thus takes *P* to be a markovian kernel rather than just a measurable function. Specifically, PAT defines the perceptual map to be a markovian kernel $P : (W \times X) \times \sigma(X) \to [0,1]$, where $\sigma(X)$ denotes the σ-algebra of *X*. Intuitively this says that the perceptual mapping assigns a probability for various perceptual states to occur given both the current state of the objective world and the current state of perception.

Given its current perception *X*, a perceptual agent (PA) has a set of actions, *G*, that it might take and must decide which action to choose. We assume that *G* is a measurable space, and define a *decision map D* to be the markovian kernel $D:(X\times G)\times\sigma(G)\rightarrow[0,1]$, where $\sigma(G)$ denotes the σ-algebra of *G*. Intuitively this says that the decision mapping assigns a probability for various actions to be selected given both the current state of perception and the last selected action.

Once an action is selected, a PA then acts on the objective world *W*. We assume that *W* has a measurable structure $\sigma(W)$, and define an *action map A* to be the markovian kernel $A:(G\times W)\times\sigma(W)\rightarrow[0,1]$. Intuitively this says that the action map assigns a probability for various states of the objective world to be selected given both the currently selected action and the current state of the objective world.

Finally, PAT assumes that there is a discrete counter, *t*, that increments each time a new perception is received. Thus $t\in\mathbb{Z}$, the integers.

Taken together, PAT defines a perceptual agent to be a six-tuple $PA=(X,G,P,D,A,t)$, where each component is as defined above.

Perceptual Agents and Bayesian Models

PAT differs from standard Bayesian decision theory (BDT) accounts of perception (e.g., Kersten, Mamassian & Yuille, 2004; Knill & Richards, 1996; Mamassian, Landy & Maloney, 2002). PAT is of course consistent with Bayes rule; the Markovian kernels *P*, *D*, and *A* can be thought of us conditional probabilities, and such kernels can extend Bayes rule to permit conditioning on sets of measure zero. Any approach to conditional probability that does not follow Bayes rule is incoherent, or pragmatically self defeating, in the sense that it is possible to make a Dutch Book against the person who fails to follow Bayes rule (de Finetti, 1937/1980; Lewis, 1980; Ramsey, 1926; Teller, 1976; but see Jeffrey, 1983).

The difference between PAT and BDT is in how Bayes rule is used and interpreted. BDT uses Bayes, and statistical decision theory more generally, to accurately estimate the true state of the world (Maloney, 2002). PAT uses Bayes in a different way: Bayesian conditional probabilities appear in the form of Markovian kernels, and are used to model the evolution of perceptual interfaces that are tuned to fitness, not to the true state of the world.

In BDT models of visual perception, we are given as input a collection of images y_0 (which is an element of a space of possible image collections *Y*) and we would like to estimate the true state *x* of the world (where *x* is an element of a space of possible states of the world *X*) . Toward this end, we would like to compute the condition probability $P(x \mid y_0)$. This conditional probability is called the *posterior* probability. By Bayes rule this posterior probability can be written $P(x \mid y_0) = P(y_0 \mid x)P(x) / P(y_0)$. The conditional probability $P(y_0 \mid x)$ is called the *likelihood* function. It can sometimes be thought of as a "rendering function," specifying how likely it is that the images y_0 would be obtained if the real state of the world were *x*. This is the sort of rendering that graphics engines do. The probability $P(x)$ is called the *prior* probability. It can be thought of as the prior assumption about the probabilities of various states *x* of the world. The probability $P(y_0)$ can be thought of as a normalizing factor that ensures that the posterior probability is in fact a probability measure.

Given one's assumptions $P(x)$ about the probabilities of states of the world, and given one's model $P(y_0 \mid x)$ of how images arise from states of the world, one can use Bayes rule to compute the posterior $P(x \mid y_0)$. The next move in the BDT approach to perception is typically to use this posterior to estimate which state *x* of the world really obtained. To do this, theorists typically use a utility function (or loss function) that specifies the benefits of getting the right answer (or the costs of getting the wrong answer). One utility function is a Dirac measure on the correct answer. This means intuitively that there is utility only in getting the right answer, and no utility for getting any other answer, even an answer close to the

correct answer. This utility function leads one to choose as the best estimate that world state *x* which has the maximum posterior probability; this choice is called the *maximum a posteriori* (MAP) *estimate.* If instead the utility falls off smoothly with the squared error, i.e, with the square of the distance to the correct answer, this leads one to choose as the best estimate that world state *x* which lies at the mean of the posterior probability (Mamassian et al., 2002). One can use other decision rules, such as maximum local mass (Brainard & Freeman, 1997) and probability matching (sampling from the posterior; e.g., Wozny, Beierholm, & Shams, 2010).

The problem with the BDT account of perception is that it requires that the space *X* of possible states *x* of the world is identical to the space of hypotheses that are entertained by the perceiver when it tries to interpret the given images y_0. That is, the BDT account of perception simply assumes that the perceiver's space of hypotheses is precisely matched to the world. This is expressed in the formalism by using the space *X* to represent both the hypotheses entertained by the perceiver and to represent the truly possible states of the world. This would of course be convenient if it were true. But the theory of evolution entails that the probability that this assumption is true is zero.

When BDT is used to model, say, the inference of 3D shape from image motion, the standard BDT interpretation is that the perceiver is recovering, or estimating, the true 3D shape in the objective world on the basis of the image motion it receives. The mapping from image motions *Y* to 3D interpretations *X*, is a map from sensory inputs to (in the normal case) true descriptions of objective reality. Evolutionary theory entails that this interpretation of the perceptual process is almost surely false.

So what different interpretation does PAT give for the inference of 3D shape from image motion? PAT says that what is happening is not an estimation of objective reality, but simply the construction of a new, more powerful, 3D perceptual interface that can guide a richer set of adaptive behaviors than the original 2D interface allowed. At some point in our evolutionary past, there must have been some fitness advantage to employing a 3D interface, an advantage that outweighed the costs in time and energy required to construct it.

Thus ITP and PAT also entail a new interpretation of psychophysical laws (Hoffman 2013). Such laws are standardly assumed by psychophysicists to describe the relationship between a subjective perceptual experience—such as the loudness of a tone or the heaviness of a weight—and the objective physical stimulus in spacetime that causes the subjective experience. But ITP denies that spacetime is, or is an aspect of, objective reality, and so it denies that there are objective physical stimuli in spacetime that cause our subjective experiences. Instead, for ITP psychophysical laws describe relationships between different aspects of our perceptual interfaces, or between different procedures for quantifying perceptual experiences.

For instance the amplitude of a sound, which is standardly assumed to be an objective property of an observer-independent physical stimulus, and its perceived loudness, which is standardly assumed to be an observer dependent subjective experience, both reside within the interface. The sound wave itself travels in spacetime, and therefore inhabits the spacetime interface of *H. sapiens*. Its amplitude, therefore, cannot be an observer-independent feature of objective reality. The psychophysical laws relating amplitude and loudness do not relate objective reality and perception, just different levels of the interface. Measurements of the acoustic level require the aid of technology, but this technology itself resides in our spacetime interface and yields reports couched in the predicates of our spacetime interface (Hoffman 2013).

Psychophysical laws do not arise because there are consistent relationships between an objective spacetime reality and our subjective experiences, but because there are consistent relationships between different aspects of our species-specific perceptual interface.

Networks of Perceptual Agents

Perceptual agents can be connected into networks of arbitrary size and complexity. The connecting mechanism is simple: The action kernel A of one agent can be all, or a component of, the perception kernel P of another agent. In this fashion, information can flow between agents within a network. Such perceptual agent networks (PANs) are computationally universal, in the sense that anything that can be computed be a universal Turing machine can also be computed by PANs. Thus PANs provide a novel architecture for creating models of learning, memory, problem solving, and other cognitive and perceptual capacities.

It is straightforward to show that any PAN is itself also a single perceptual agent (PA). Thus PANs provide a flexible and powerful "object-oriented" framework for cognitive and perceptual modeling. One can, for instance, create a PAN that solves the *motion-correspondence* (MC) problem: It takes as input a discrete sequence of video frames in which each frame contains a few dots, and it decides which dots are the same between frames and how they move in two dimensions between frames. This complex MC PAN can then be written as a single MC PA — a move one can call "coarse graining." One can then create a PAN that solves the *structure-from-motion* (SFM) problem: It takes as input the two-dimensional motions of dots constructed by the MC PAN, then decides if these motions are consistent with being the perspective projections of a three-dimensional structure moving rigidly in space, and, where possible, computes the relevant three-dimensional structures and motions. This SFM PAN can be coarse grained into a single SFM PA. Then the MC PA and the SFM PA can be coarse grained together to create a single MC-SFM PA. In this fashion one can use the technology of PANs to flexibly model cognitive and perceptual processes, to coarse grain them when convenient, and unpack them when necessary.

There is a simplest PA, viz., an agent whose measurable spaces X and G each have just two states, allowing them each to be represented by just 1 bit of information, and whose kernels are the simplest maps between these spaces. Thus one can build PAN circuits from the ground up, starting with the simplest 1 bit PAs. Once an elementary circuit has been constructed, it can itself be considered a single PA, i.e., it can be coarse grained. The PAs obtained by coarse graining can themselves be used to construct new, more sophisticated PANs. These new PANs can themselves be coarse grained and used to construct even more sophisticated PANs, ad infinitum.

The key to this computational power of PAs and PANs is the computing potential of the kernels D, A and P. Consider, for instance, the case where X and G each are one bit, and the kernel

$$D : (X_t \times G_t) \times \sigma(G_{t+1}) \to [0,1]$$

flips the state of the G_t bit if the X_t bit is 1, and leaves the state of the G_t bit unchanged if the X_t bit is 0. This is the well-known controlled-NOT (cNOT) operator. We can construct a matrix representation of the Markovian kernel *D* as follows. We write the X_t bit and G_t bit side by side, with the X_t bit to the left. So, 00 means that both X_t and G_t are 0; 01 means X_t is 0 and G_t is 1; 10 means X_t is 1 and G_t is 0; 11 means X_t is 1 and G_t is 1. We can name these four states by their binary numbers: 00 is state 0; 01 is state 1; 10 is state 2; and 11 is state 3. We then think of these four states as orthogonal vectors in a four-dimensional space. Each orthogonal vector has a 1 in the position corresponding to its binary number, and zeroes elsewhere. So 00 becomes (1,0,0,0); 01 becomes (0,1,0,0); 10 becomes (0,0,1,0); 11 becomes (0,0,0,1). Then in this basis the markovian kernel *D* can be written as the stochastic matrix

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

We can check that this matrix does what we want. If X_t is 0, then we want the G_t bit to remain unchanged. Thus we want *D* to leave the states 00 = (1,0,0,0) and 01 = (0,1,0,0) unchanged, which it does. If X_t is 1, then we want to flip the G_t bit. Thus we want *D* to swap the states 10 = (0,0,1,0) and 11 = (0,0,0,1), which it also does.

In similar fashion we can construct Markovian kernels for all the other computational operations we want. We can then construct PAs and PANs, using these computational operations, to model any aspects of perception and cognition that we wish. The PA formalism and PA networks will, I suspect, prove to be a more flexible and powerful framework for such modeling than neural networks, in part because any PAN can be coarse grained to a single PA.

Discussion

ITP is such a departure from standard theories of perception, that it is easily misunderstood. For instance, one might think that when ITP says that we don't see objective reality as it is, that what ITP is really claiming is that we do see some, but not *all*, of reality as it is. We see only those portions of reality that we need to see, the portions that are necessary for enhancing our fitness and our chances of survival. On this interpretation, ITP is not much different from well-known ecological approaches to perception, in which we perceive those aspects of reality that afford the information we need for survival, and we don't perceive much else. We see, on this account, light with wavelengths between roughly 400 and 700 nanometers because this is the part of reality that we need to see to survive and reproduce. We see the real surfaces of physical objects — but not atoms and quarks — because in our niche this is the aspect of reality we need most to know in order to survive.

But ITP says something far more radical. The Fitness-Beats-Truth Theorem entails that the probability is one that the very predicates of our perceptions — space, time, objects, colors, motions, pitch, tembre, tastes, smells — are simply the wrong predicates to describe reality as it is. *None* of our perceptions are veridical. All of our perceptions have evolved to guide adaptive behaviors, but this evolution did not result in us seeing parts of reality truly. We see *none* of reality truly. Our perceptions are simply a species-specific interface that we must take seriously, because it has evolved to help us survive and reproduce. But none of it is literally true.

This claim of ITP is radical, but it is nevertheless an empirical claim. It makes falsifiable predictions. It predicts, as we discussed above, that space-time is doomed. Space and time are not fundamental features of objective reality. The predicates of space and time are predicates that a particular species has evolved as a useful shorthand to guide adaptive behaviors. They are not an insight. Physicists will discover that spacetime is not the right language or framework for a theory of reality as it is. They will discover that theories that are forced to be stated in the language of spacetime will miss deep symmetries of nature, and will probably also be unnecessarily complex. They will discover that there is a deeper theory of reality, one that is non spatial and non temporal, from which spacetime can be derived as a consequence. An example of this kind of theory is the work of Seth Lloyd (2006), in which he posits that quantum information and

quantum computations — not in space and time but just in themselves — might be the fundamental reality. He then shows how discrete patches of spacetime might be associated to each quantum gate, with the curvature of a spacetime patch being related to the amount of action at the gate. Combining these patches together can give rise to spacetime and a theory of quantum gravity. On this approach, spacetime is not fundamental, but derivative upon a more fundamental reality that is beyond spacetime. Lloyd's specific proposal might, of course, not be right. But it gives a concrete idea of the kind of proposal that ITP predicts must eventually be found — a proposal in which spacetime is explicitly derived from a different reality that is far more fundamental.

ITP predicts, as we discussed above, that no physical object in spacetime has definite values of any dynamical properties when it is not observed. No electron has a position, no atom has a momentum, when it is not observed. All careful tests of this prediction find violations of Bell's inequalities, a result that is compatible with, indeed predicted by, ITP.

ITP predicts that neurons, being just one kind of physical object in spacetime, do not exist when they are not observed. This entails that neurons cause none of our behaviors, none of our mental states, none of our conscious experiences: Something that does not exist when it is not observed cannot be a *fundamental* cause of anything.

Thus ITP explains why the hard problem of consciousness has remained intractable for centuries. The hard problem is to explain how conscious experiences, such as the experience of the taste of chocolate, can be caused by or can somehow arise from purely physical systems whose components are, by hypothesis, not conscious, i.e., are not themselves subjects of experience. The hard problem arises for the simplest kinds of conscious experiences, even in the absence of more sophisticated forms of consciousness, such as self awareness. According to ITP and the FBT Theorem, the reason that the hard problem has not been solved is that it cannot be solved. The formulation of the hard problem has assumed that objects in spacetime have genuine causal powers. The FBT Theorem entails that they almost surely do not. Trying to build a causal account of the provenance of consciousness starting with components that have no causal powers is an exercise in futility. One cannot explain how neural activity causes conscious experiences if neural activity is not the kind of thing that can be the source of causes.

ITP and FBT point to another reason for the intractability of the hard problem: The language of physical objects in spacetime is simply the wrong language for stating a theory of consciousness and its provenance. It is the wrong language, because it evolved for a different purpose — to keep us alive and guide adaptive behavior. For that purpose it is quite effective. It did not evolve for the purpose of stating true theories about the nature of consciousness and reality more generally, and it is completely unsuited to the task. It is the language of the perceptual interface, not the language needed to describe the reality that, in the best interests of the organism, is *hidden* from view by that interface. Requiring a scientist to devise a theory of consciousness using the language of physical objects in spacetime is like requiring a student to give a theory of the operation of a computer's microprocessor using the language of the pixels of the desktop interface. No matter how smart the student might be, the game is rigged against their success, because they have been saddled with a language that cannot succeed in solving the problem. This is the problem that has stopped any progress in physicalist attempts to solve the hard problem of consciousness. ITP does not, by itself, propose a solution to the hard problem (for such a proposal see Hoffman and Prakash, 2014). It simply gives a diagnosis for why the problem has so long remained incurable.

Is ITP a brand new idea? No. There is a long history of similar ideas, a history that is helpfully discussed by Koenderink (2015) and Mausfeld (2015).

In particular, although exegesis of Kant is notoriously controversial, his distinction between noumena and phenomena is roughly similar to the distinction in ITP between objective reality and a species-specific perceptual interface. The claim of ITP about spacetime—that it is not an aspect of objective reality but is instead simply analogous to the desktop of our perceptual interface—roughly parallels Kant's claim in the *Critique of Pure Reason* (1781/1922 p. 21) that "It is, therefore, solely from the human standpoint that we can speak of space, of extended things, etc. If we depart from the subjective condition under which alone we can have outer intuition, namely, liability to be affected by objects, the representation of space stands for nothing whatsoever." But whereas Kant comes to this conclusion through his controversial notion of the synthetic a priori, ITP comes to it as a theorem that follows from evolutionary game theory. And if Kant is correctly interpreted as claiming that a science of the noumena is not possible, then ITP is free to disagree.

ITP itself offers no such a science. But the theory of conscious agents (Hoffman and Prakash, 2014) is one proposal towards such a science that is consistent with ITP.

ITP agrees in part with the model-dependent realism of Hawking and Mlodinow (2010) when they say "There is no way to remove the observer— us— from our perception of the world, which is created through our sensory processing and through the way we think and reason. Our perception— and hence the observations upon which our theories are based— is not direct, but rather is shaped by a kind of lens…" (p. 46). Indeed according to ITP we do not see, directly or indirectly, objective reality; we see certain aspects of relevant fitness functions formatted according to a species-specific interface.

But ITP disagrees with how Hawking and Mlodinow end that last sentence: "…is shaped by a kind of lens, the interpretive structure of our human brains." Here they appear to attribute causal power to the brain, an attribution more clearly suggested a few sentences later: "The brain, in other words, builds a mental picture or model" (p. 47). If they do intend to attribute causal power to the brain, then ITP disagrees: Although it is a harmless, and even useful, fiction for most research in neuroscience to think of neural activity as causing behavior and mental states, it is nonetheless a fiction, and one that is no longer benign if the topic of research is the mind-body problem.

One can deny that brains have causal powers without flouting the fundamental tenet of model-dependent realism, viz., "… it is pointless to ask whether a model is real, only whether it agrees with observation. If there are two models that both agree with observation … then one cannot say that one is more real than another" (p. 46). Here is the key fact: There simply is no model of the brain that explains, without invoking magic at the crucial point, how brains cause conscious experiences. To play the game of model-dependent realism—for a brain model to be just as real as any other model—one first must have a genuine model.

Although there is a long history of ideas similar to ITP, nevertheless the FBT Theorem and evolutionary games on which ITP relies for its sweeping conclusions are new. Moreover, the specific metaphor of the desktop interface, with spacetime playing the role of the desktop and physical objects the role of icons on the desktop, appears to be new. But, whether new or not, ITP, with its rejection of an objective spacetime

reality, parts company with all major contemporary theories of perception. If ITP is correct, then even their basic accounts of illusions, hallucinations and psychophysical laws will not survive.

REFERENCES

Allen, J. A., & Clarke, B. C. (1984). Frequency-dependent selection— homage to Poulton, E.B. *Biological Journal of the Linnean Society*, 23, 15–18.

Ansmann, M., Wang, H., Bialczak, R. C., Hofheinz, M., Lucero, E., Neeley, M., . . . Martinis, J. M. (2009). Violation of Bell's inequality in Josephson phase qubits. *Nature*, 461, 504–506. doi:10.1038/ nature08363

Barnard, C. J., & Sibly, R. M. (1981). Producers and scroungers: A general model and its application to captive flocks of house sparrows. *Animal Behavior*, 29, 543–550.

Bomze, I. M. (1983). Lotka-Volterra equations and replicator dynamics: A two dimensional classification. *Biological Cybernetics*, 48, 201– 211.

Brewer, W. (2011). *Perception and its objects*. New York: Oxford, UK: Oxford University Press.

Byrne, A., & Hilbert, D. R. (2003). Color realism and color science. *Behavioral and Brain Sciences*, 26, 3–26.

Cabello, A., Estebaranz, J. M., Garc´ıa-Alcaine, G. (1996). Bell-Kochen-Specker theorem: A proof with 18 vectors," Physics Letters, A, 212, 183.

Campbell,J., & Cassam,Q. (2014). Berkeley's puzzle: what does experience teach us? Oxford, UK: Oxford University Press.

Chater, N., & Vitányi, P. M. B. (2003). The generalized universal law of generalization. *Journal of Mathematical Psychology*, 47, 346-369.

Coates, P. (2007). The metaphysics of perception. New York: Routledge.

Dawkins, R. (1983). Universal darwinism. In Evolution from molecules to man, ed. D. S. Bendall. New York: Cambridge University Press.

de Finetti, B. (1937). La Prevision: ses lois logiques, se sources subjectives. *Annales de l'Institut Henri Poincare*, 7: 1–68; Translated into English and reprinted in Kyburg and Smokler, *Studies in Subjective Probability*, Huntington, NY: Krieger, 1980.

Dennett, D. C. (2005). Darwin's dangerous idea. New York: Touchstone Press.

Fields, C. (2014). This boundary-less world. In *Brain, mind, cosmos*, ed. D. Chopra. La Costa, CA: Chopra. Chapter 13.

Fish, W. (2009). Perception, hallucination, and illusion. New York: Oxford University Press.

Fish, W. (2010). Philosophy of perception: a contemporary introduction. New York: Routledge.

Fodor, J.A. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

Fuchs, C. (2010). QBism, the perimeter of quantum Bayesianism. arXiv:1003.5209v1.

Giddings, S. (2015). Spacetime. In J. Brockman (Ed), *This idea must die*. New York: Harper Perennial.

Giustina, M., Mech, A., Ramelow, S., Wittmann, B., Kofler, J., Beyer, J., Lita, A., Calkins, B., Gerrits, T., Nam, S.W., Ursin, R., Zeilinger, A., (2013-04-14). Bell violation using entangled photons without the fair-sampling assumption. Nature, 497, (7448): 227- 30.

Harper, G. R., & Pfennig, D. W. (2007). Mimicry on the edge: Why do mimics vary in resemblance to their model in different parts of their geographical range? *Proceedings of the Royal Society B*, 274, 1955– 1961.

Hawking, S., & Mlodinow, L. (2010.) *The grand design*. New York: Bantam.

Hofbauer J., & Sigmund K. (1998). Evolutionary games and population dynamics. New York: Cambridge University Press.

Hoffman, D. D. (1998). *Visual intelligence: how we create what we see*. New York: W. W. Norton.

Hoffman, D.D. (2008). Conscious realism and the mind-body problem. *Mind & Matter, 6,* 87– 121.

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

Hoffman, D. D. (2011). The construction of visual reality. In J. Blom & I. Sommer (Eds.) *Hallucinations: theory and practice*, New York: Springer, pp. 7–15.

Hoffman, D. D. (2012). The sensory desktop. In J. Brockman (Ed.) *This will make you smarter:* new scientific concepts to improve your thinking. New York: Harper Perennial, pp. 135– 138.

Hoffman, D. D. (2013). Public objects and private qualia: the scope and limits of psychophysics. In L. Albertazzi (Ed.) *The Wiley-Blackwell handbook of experimental phenomenology*. New York: Wiley-Blackwell, pp. 71–89.

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

Hoffman, D. D., & Singh, M. (2012). Computational evolutionary perception. *Perception, 41*, 1073–1091.

Hoffman, D.D., Singh, M., Mark, J. (2013). Does evolution favor true perceptions? Proceedings of the SPIE 8651, Human Vision and Electronic Imaging XVIII, 865104. DOI: 10.1117/12.2011609.

Hoffman, D. D., Singh, M., Prakash, C. (2015a). The interface theory of perception. *Psychonomic Bulletin and Review,* 22, 1480-1506. DOI 10.3758/s13423-015-0890-8.

Hoffman, D. D., Singh, M., Prakash, C. (2015b). Probing the interface theory of perception; Replies to commentaries. *Psychonomic Bulletin and Review*, 22, 1551-1576. DOI 10.3758/s13423-015-0931-3.

Jeffrey, R. (1983). *The Logic of Decision*, 2nd ed., Chicago: University of Chicago Press.

Kant, I. (1781/1922). Critique of pure reason. F.M. Müller (trans.), Second Edition. New York: Macmillan

Kersten, D., Mamassian, P., Yuille, A. L. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 555, 271–304.

Knill, D., Richards, W. A. (Eds.). (1996). *Perception as Bayesian inference*. New York: Cambridge University Press.

Koenderink, J.J. (2011). Vision as a user interface. *Human Vision and Electronic Imaging XVI, SPIE Vol. 7865*. doi: 10.1117/12.881671.

Koenderink, J.J. (2013). World, environment, umwelt, and inner-world: a biological perspective on visual awareness, *Human Vision and Electronic Imaging XVIII, SPIE Vol. 8651*. doi: 10.1117/12.2011874.

Koenderink, J.J. (2014). The all seeing eye? *Perception*, *43*, 1–6.

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

Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In Richard C. Jeffrey (ed.), *Studies in Inductive Logic and Probability* (Vol. 2), Berkeley: University of California Press, 263–293.

Lieberman, E., Hauert, C., & Nowak, M. A. (2005). Evolutionary dynamics on graphs, *Nature*, 433(7023), 312–316.

Lloyd, S. (2006). *Programming the universe*. New York: Knopf.

Mamassian, P., Landy, M., Maloney, L. T. (2002). Bayesian modeling of visual perception. In R. Rao, B. Olshausen, & M. Lewicki (Eds.), *Probabilistic models of the brain: perception and neural function* (pp. 13–36). Cambridge, MA: MIT Press.

Maloney, L. (2002). Statistical decision theory and biological vision. In D. Heyer and R. Mausfeld (Eds) Perception and the physical world: Psychological and philosophical issues in perception (pp. 145 - 188). New York: Wiley.

Mark, J.T. (2013). Evolutionary pressures on perception: when does natural selection favor truth? Ph.D. Dissertation, University of California, Irvine.

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

Marion, B.B. (2013). The impact of utility on the evolution of perceptions, Ph.D. Dissertation, University of California, Irvine.

Mausfeld, R. (2002). The physicalist trap in perception theory. In D. Heyer, & R. Mausfeld (Eds.) *Perception and the physical world: psychological and philosophical issues in perception*. New York: Wiley, pp.75–112.

Mausfeld, R. (2015). Notions such as “truth” or “correspondence to the objective world” play no role in explanatory accounts of perception. *Psychonomic Bulletin and Review,* 22, 1535-1540.

Mermin, N.D. (1985). Is the moon there when nobody looks? Reality and the quantum theory. *Physics Today*, April, 38-47.

Mitchell, M. (1998). *An introduction to genetic algorithms*. Cambridge, MA: Bradford Books

MIT Press.

Nowak, M. (2006). *Evolutionary dynamics: exploring the equations of life*. Cambridge, MA: Belknap Press of Harvard University Press.

Pan, J.-W., Bouwmeester, D., Daniell, M., Weinfurter, H., Zeilinger, A. (2000). Experimental test of quantum nonlocality in three-photon GHZ entanglement. *Nature*, 403, (6769): 515–519.

Pinker, S. (2005). So how does the mind work? *Mind & Language*, *20,* 1-24.

Pizlo, Z. (2012). 3D shape: Its unique place in visual perception. Cambridge, MA: MIT Press.

Pizlo, Z. (2015). Philosophizing cannot substitute for experimentation: comment on Hoffman, Singh & Prakash. *Psychonomic Bulletin and Review,* 22, 1546-1547. DOI 10.3758/s13423-014-0760-9.

Ramsey, F. P. (1926). Truth and Probability. In Richard B. Braithwaite (ed.), *Foundations of Mathematics and Other Logical Essay*, London: Routledge and Kegan Paul, 1931, pp. 156–198.

Rowe, M.A., Kielpinski, D., Meyer, V., Sackett, C.A., Itano, W.M., Monroe, C., Wineland, D.J. (2001). Experimental violation of a Bell's inequality with efficient detection. Nature, 409, (6822): 791–94.

Salart, D., Baas, A., van Houwelingen, J. A. W., Gisin, N., Zbinden, H. (2008). Spacelike separation in a Bell test assuming gravitationally induced collapses, *Physical Review Letters*, 100, (22): 220404.

Samuelson, L. (1997). Evolutionary games and equilibrium selection. Cambridge, MA: MIT Press.

Sandholm, W. H. (2007). Population games and evolutionary dynamics. Cambridge, MA: MIT Press.

Searle, J. (2015). Seeing things as they are: a theory of perception. New York: Oxford University Press.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science* 237, 1317–1323.

Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin and Review,* 1, 2–28.

Shepard, R. N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral and Brain Sciences*, 24, 58-601.

Singh, M., & Hoffman, D. D. (2013). Natural selection and shape perception: shape as an effective code for fitness. In S. Dickinson and Z. Pizlo (Eds.) *Shape perception in human and computer vision: an interdisciplinary perspective*. New York: Springer, pp. 171–185.

Taylor, P., & Jonker, L. (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40, 145–156.

Teller, P. (1976). Conditionalization, Observation, and Change of Preference. In W. Harper and C.A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Dordrecht: D. Reidel.

Tenenbaum, J. B. and T. L. Griffiths (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences* 24, 629–641.

von Sydow, M. (2012). From Darwinian metaphysics towards understanding the evolution of evolutionary mechanisms. A historical and philosophical analysis of gene-Darwinism and universal Darwinism. Universitätsverlag Göttingen.

Weihs, G., Jennewein, T., Simon, C., Weinfurter, H., Zeilinger, A. (1998). Violation of Bell's inequality under strict Einstein locality conditions, *Physical Review Letters*, 81, 5039.

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology*, 6(8), e1000871.