Psychon Bull Rev (2015) 22:1551–1576
DOI 10.3758/s13423-015-0931-3

Probing the interface theory of perception: Reply to commentaries

Donald D. Hoffman[1] **· Manish Singh**[2] **· Chetan Prakash**[3]

Published online: 30 September 2015

Abstract We propose that selection favors nonveridical perceptions that are tuned to fitness. Current textbooks assert, to the contrary, that perception is useful because, in the normal case, it is veridical. Intuition, both lay and expert, clearly sides with the textbooks. We thus expected that some commentators would reject our proposal and provide counterarguments that could stimulate a productive debate. We are pleased that several commentators did indeed rise to the occasion and have argued against our proposal. We are also pleased that several others found our proposal worth exploring and have offered ways to test it, develop it, and link it more deeply to the history of ideas in the science and philosophy of perception. To both groups of commentators: thank you. Point and counterpoint, backed by data and theory, is the essence of science. We hope that the exchange recorded here will advance the scientific understanding of perception and its evolution. In what follows, we respond to the commentaries in alphabetical order.

Keywords Bayesian inference · Parameter estimation · Perceptual categorization and identification · Visual perception · Categorization

✉ Donald D. Hoffman
ddhoff@uci.edu

1 Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA

2 Department of Psychology and Center for Cognitive Science, Rutgers University, New Brunswick, NJ, USA

3 Department of Mathematics, California State University, San Bernardino, CA, USA

Barton Anderson
Where does fitness fit in theories of perception? doi:10.3758/s13423-014-0748-5

Overview (1) Anderson argues that, "For the games they considered, the only 'force' of adaptation was through natural selection. In such cases, perception should indeed track utility directly. But the same doesn't hold if animals can adjust their behavior to meet homeostatic demands on ontogenetic time scales." We reply that time scale and ontogenetics are irrelevant to the issue of tracking utility. (2) Anderson says, "If the 'payoff' function of homeostasis is nonmonotonic, as is typical, then the perceptual response needs to track resources monotonically so that an animal can know to adapt its behavior to achieve homeostasis." We explain why this is false. (3) Anderson argues that our characterization of perceptual strategies using an abstract set of world states W assumes a "God's eye view" of the world and "misses a fundamental point about what constitutes both the objects of science and experience; the fundamental elements of system description in science are observables, not unspecifiable 'world states'. " We explain how our abstract approach serves to avoid a God's eye view, and how it engages the standard interaction of theory, observables, and experiments that is central to science. (4) Anderson claims that our argument for redefining the notion of perceptual illusion is that, "Since we have no access to the 'true' world states, it is impossible to ever determine whether perception is veridical (in their sense) or in error, which leads them to the conclusion that such distinctions can only be defined in terms of adaptive behavior." We explain that this is not our argument, and then explain what our real argument is. (5) Anderson proposes to define veridical perception in terms of the coherence between different observables. We explain why this is misleading and inadequate. (6) Anderson argues that, "The striking metaphysical claim of IT [interface theory] is that our experience, and all other means of measuring the world, provide no information about the world 'as it is,' but merely reveal properties of an adaptive interface. So what would change if this were true? . . . This generates skepticism about the ultimate veracity of the properties we ascribe to physical reality, but is otherwise seemingly inconsequential." We explain precisely why it is consequential, that it has repeatedly been tested by careful experiments, and that the results of each experiment are precisely as predicted by the interface theory. (7) Anderson argues that the interface theory of perception "flies in the face of experience; there are many perceptual scales that vary monotonically with physical resources, including the sensory laws that spawned the birth of psychology as a natural science." We explain why the interface theory predicts such perceptual scales. (8) Anderson notes that the philosophical school of pragmatism, going back at least to John Dewey, makes claims similar to those of the interface theory. We agree.

We now consider these points in more detail.

(1) Anderson argues that, "For the games they considered, the only 'force' of adaptation was through natural selection. In such cases, perception should indeed track utility directly. But the same doesn't hold if animals can adjust their behavior to meet homeostatic demands on ontogenetic time scales."

Anderson grants us that on phylogenetic time scales the learning algorithm called natural selection will generate perceptual systems that track utility directly. But he then claims that the learning systems that adjust behavior on ontogenetic time scales will necessarily generate *different* kinds of perceptual systems—ones that track truth rather than utility.

How shall we interpret this claim? Perhaps the claim is that some theorem about learning systems entails that what can be learned is in principle a function of the *time scale* of learning: What can be learned on long time scales is in principle different in critical ways from what can be learned on shorter time scales. Anderson cites no such theorem. We know of no such theorem, and doubt that one exists. Or perhaps the claim is not about time scale but about the nature of *ontogenetic* learning, and the claim is that ontogenetic learning is not shaped, or is less shaped, by fitness payoffs than is phylogenetic learning. Again, we know of no evidence for such a claim, and it seems implausible.

Another way to try to interpret Anderson's claim is as an empirical one: That the basic perceptual systems an organism is born with do track utility, but elaborations of these perceptual systems are required ontogenetically in order to maintain homeostasis, and these elaborations necessarily involve perceptions that track truth.

But this interpretation is flatly contradicted by superprecocial species such as Megapode birds. They enter life able to live independently from day one. Thus, using their innate perceptual systems that were shaped by natural selection to track utility, they are, without further learning, immediately able to act in ways that maintain homeostasis.

Thus, either as a theoretical claim about the nature of learning or an empirical claim about the perceptions of organisms, Anderson's argument about short time scales has no traction.

(2) Anderson says, "If the 'payoff' function of homeostasis is nonmonotonic, as is typical, then the perceptual response needs to track resources monotonically so that an animal can know how to adapt its behavior to achieve homeostasis."

Anderson's idea here is that one needs to know, for example, whether one is above or below the homeostatic ideal so that one knows how to act to restore homeostasis. (Here we assume that for Anderson the word "know" is not restricted to *conscious* knowledge. Our reply uses "know" in this unrestricted sense that includes unconscious knowledge.) As he puts it, "An interface strategy of the type advocated in Fig. 3 provides no information about what an organism should do to achieve homeostasis; it makes the problem of homeostasis (or cooking) a complete guessing game. Do I need more or less salt to achieve homeostasis? Is there too little or too much salt in this dish?"

How shall we interpret this claim? Perhaps he is claiming it is logically impossible for a system to maintain homeostasis if its perceptual systems only track fitness payoffs and not the true state of the world. But this claim is surely false. If an organism has a collection of actions at its disposal it can, for example, find the *gradient* of fitness payoffs as a function of the various actions and then act so as to increase its payoff. All it needs to know are the fitness consequences of actions—does this act increase or decrease fitness? It does not need to know or track the true state of the world. In the theory of reinforcement learning, for instance, an agent can have "partial observability," in which it only knows the reward consequences of its actions. This is distinguished from the case of "full observability," in which the agent knows both the reward and the state of the world. Partial observability does not preclude acting so as to optimize fitness or maintain homeostasis.

Moreover, in many cases of practical interest, the fitness function depends on not just one variable in the world, but many. For instance, the fitness payoff obtained from eating an item of food depends on how much of each amino acid it contains, how much sugar and salt it has, its profile of a long list of vitamins, its profile of a long list of minerals, its profile of a long list of polyphenols, and so on. In such a case, it is not feasible to track the true value of every variable, but it is feasible to learn the fitness consequences of acts such as eating an egg versus eating poison ivy, even in complete ignorance of the very existence of vitamins, minerals, polyphenols, and amino acids.

So Anderson's claim here is a logical non sequitur, and empirically false.

(3) Anderson argues that our characterization of perceptual strategies using an abstract set of world states W assumes a "God's eye view" of the world and "misses a fundamental point

about what constitutes both the objects of science and experience; the fundamental elements of system description in science are observables, not unspecifiable 'world states.' "

This argument gets things exactly backward. Suppose that instead of letting *W* be an arbitrary set whose properties we do not pretend to know, we instead assert that *W* is, say, a specific four-dimensional space with a specific Minkowski metric that is seen with a specific perceptual strategy. Then, indeed, we would be guilty of taking a God's eye view: How in the world could we know that *W* has that specific structure, and that we use that specific perceptual strategy, unless we were God?

It is precisely because we do not have a God's eye view that we allow *W* to have *any* structure a priori and then simply classify *all* of the logical possibilities for perceptual strategies. This allows us to take an evolutionary point of view, rather than a God's eye view, and use evolutionary game theory (rather than God) to determine which perceptual strategies are favored by natural selection. In the process, we learn that our perceptions almost surely do not track the structure of *W*, which entails that some assumptions that we naturally make about *W*—such as that it has three dimensions of space, a dimension of time, and contains physical objects with properties such as mass and position—are almost surely false. This means we have discovered that we know *less* about the objective world then we previously assumed. This is just the opposite of the all-knowing God's eye view.

According to Anderson, not specifying a priori what is the structure of *W* "misses a fundamental point about what constitutes both the objects of science and experience; the fundamental elements of system description in science are observables, not unspecifiable 'world states.' "

Not at all. If one wants to understand scientifically which perceptual strategies are in principle favored by natural selection, one cannot assume that one knows a priori what *W* is. One must instead pose the harder question: Is any perceptual strategy favored in *almost all* possible worlds *W*? This is a precise technical question, which precisely requires us not to bias the outcome by an a priori assumption about the nature of the world states of *W*. We discover that natural selection favors certain interface strategies tuned to fitness. We also discover the invention-of-symmetry theorem, which strictly limits what we can infer about *W* from the invariance properties of our perceptions and actions. Sometimes the progress of science reveals limits about what we can know or do, as in Heisenberg's uncertainty principle or Gödel's incompleteness theorems. The interface theory reveals that we know far less about reality as it is than we naively thought we knew.

(4) Anderson claims that our argument for redefining the notion of perceptual illusion is that, "Since we have no access to the 'true' world states, it is impossible to ever determine whether perception is veridical (in their sense) or in error, which leads them to the conclusion that such distinctions can only be defined in terms of adaptive behavior."

This is not our argument. Instead, our argument begins by recognizing that the theory of evolution entails that our perceptions are almost surely not veridical. Thus, to define illusions as perceptions that are not veridical, which has been the standard definition of illusion, would require us to take all perceptions as illusions—not very helpful. But we think that a useful distinction *can* be made between illusory and nonillusory perceptions. The natural distinction in an evolutionary framework is between perceptions that guide adaptive behaviors and those that do not. It is this line of argument that led us to our redefinition of illusion.

(5) Anderson proposes to define veridical perception in terms of the coherence between different observables.

This is inadequate and misleading. We already have the term "coherence" to describe the agreement among observables. It would only lead to confusion to try to take the term "veridical," which already has its own well-entrenched connotations, and try to force it to mean coherence. The word "veridical," with its present meaning, has an important role to play in perceptual theory. It stands for an important theory about the relationship between our perceptions and the world. That theory, evolution tells us, is almost surely false. But it is nevertheless an important theory that deserves to keep its name.

(6) Anderson argues that "The striking metaphysical claim of IT is that our experience, and all other means of measuring the world, provide no information about the world 'as it is,' but merely reveal properties of an adaptive interface. So what would change if this were true? . . . This generates skepticism about the ultimate veracity of the properties we ascribe to physical reality, but is otherwise seemingly inconsequential."

Here is a consequence: Interface theory predicts that *no* physical object has definite values of *any* dynamical physical property when it is not observed. For instance, no electron has any definite position or spin when it is not observed.

This is a clean prediction of the theory. If it is proves to be empirically false, then Interface theory is wrong. If it is empirically supported, then the standard realist view of most perception scientists is wrong.

One might object that this is hardly a real prediction, because it is in principle untestable. How in the world could one empirically test whether, say, an electron has a definite position or spin *when it is not being observed*? This seems no better than asking how many angels can dance on the head of a pin.

But it *is* empirically testable, it has already been subjected repeatedly to careful experiments, and the outcome of *every* experiment conducted so far has been *entirely as predicted by Interface theory*: No physical object has definite values of any dynamical physical property when it is not observed. The literature on this remarkable sequence of experiments is extensive, but see, for example, Ansmann et al. (2009); Cabello, Estebaranz, and García-Alcaine (1996); Fuchs (2010); Giustina et al. (2013); Pan, Bouwmeester, Daniell, Weinfurter, and Zeilinger (2000); Rowe et al. (2001); Salart, Baas, van Houwelingen, Gisin, and Zbinden (2008); and Weihs, Jennewein, Simon, Weinfurter, and Zeilinger (1998). A gentle introduction to this literature is given by Mermin (1985).

Anderson closes his commentary by saying, “remarkable theories require remarkable evidence, and there is currently insufficient evidence to take the metaphysical leap proposed by interface theory.” The literature just cited is precisely the remarkable evidence that is required.

(7) Anderson argues that the interface theory of perception, “flies in the face of experience; there are many perceptual scales that vary monotonically with physical resources, including the sensory laws that spawned the birth of psychology as a natural science.”

According to the interface theory, the reason that many sensory laws vary monotonically with physical properties is that the measured world of physics is simply an extension, via symmetry groups, of basic predicates of the human perceptual interface, such as space and time. A mistake in standard psychophysics, going back to Fechner, is to think that the measured world of physics is the objective world as it really is, and that therefore psychophysical laws are relationships between objective reality and subjective experience. They are not. They are relationships between different levels of description of our perceptual interface.

Moreover, as we mention in Point 6 above, the results of careful experiments in physics are entirely compatible with, and in fact predicted by, the hypothesis that the fundamental dynamical properties of physics—including position, momentum, and spin—do not describe reality as it is, but are instead products of—that is, creations of—the measurement process. As quantum Bayesianism (also known as *QBism*) puts it, “That the world should violate Bell’s theorem remains, even for QBism, the deepest statement ever drawn from quantum theory. It says that quantum measurements are moments of creation” (Fuchs, 2010, p. 14).

(8) Anderson notes that the philosophical school of pragmatism, going back at least to John Dewey, makes claims similar to those of the interface theory.

We agree. We would point out, however, that among our new contributions are a mathematically rigorous use of evolutionary game theory to provide scientific evidence for pragmatism, our proof of the invention-of-symmetry theorem that eviscerates a standard objection to pragmatism, and the construction of a mathematically precise interface theory of perception that captures key insights of pragmatism.

Jonathan Cohen

Perceptual representation, veridicality, and the interface theory of perception. doi:10.3758/s13423-014-0782-3

Overview (1) Cohen states that our definition of veridical perception requires an absurd identity theory in which x veridically represents y iff $x = y$. We reply that this is a misunderstanding of our text that is easily set straight; we do not require the identity $x = y$. We simply require a homomorphism. (2) Cohen argues that it is obvious that we must specify the contents of perceptions before we can evaluate them for truth or falsity. We reply that it is obvious that we do not need this. In formal logic, we do not need to specify the contents of p and q before we can conclude that $p \vee q \Rightarrow p$ is a fallacy or that $p \wedge (\neg p \vee q)$ is false. In information theory, we do not need to specify the contents of messages before we can correct errors in their transmission. Similarly, in the logic of evolutionary games, we do not need to specify the contents of perceptions before we can evaluate them for truth or falsity. The power of logic, information theory, and evolutionary game theory is to arrive at truths that are independent of assignments of content. (3) Cohen proposes specific content assignments for the jewel beetle and dragonfly examples so that their perceptions turn out to be veridical, and then claims that we are not entitled to make content assignments on which they turn out to be nonveridical. We reply that our evolutionary games, which were formulated with no a priori assignment of contents to perceptions, do allow us to rule out the content assignments that Cohen proposes, and moreover to rule out all theories of perceptual content currently proposed by philosophers of perception. (4) Cohen argues that our evolutionary games show that perception veridically represents fitness payoffs. We reply that fitness payoffs, unlike states of the world, are not organism-independent features of the world: They cease to exist if the organism ceases to exist. Perceptual experiences are satisficing solutions to the problem of getting more fitness points than your competitors; they are not veridical representations of fitness functions or of objective features of the world.

We now consider these points in more detail.

(1) Cohen argues that we incorrectly define veridical perception. In particular, he argues that we define veridical perception in a way that “commits the proponent of veridical representation to something deeply implausible, and so stacks the deck against such views before we even get to the arguments. The condition that HSP [in our original article] here take as defining veridical perception—that $X \subseteq W$—amounts to an absurd identity theory, on which x veridically represents y iff $x = y$. But, exactly because such an identity theory is absurd, it is an unreasonable assumption for HSP to foist on their opponents.”

We agree that the condition $X \subseteq W$ is not necessary for veridical perception, and we understand that many proponents of veridical perception do not accept this condition. What they often do accept is that the perceptual map $P : W \rightarrow X$ is a homomorphism that preserves all structures on W, which is the condition that characterizes critical realist strategies (Definition 5).

So let us be quite clear about our position: Critical realist strategies, as specified in Definition 5 and in which $X \not\subset W$, are strategies of veridical perception. We have tested these strategies in a variety of evolutionary games and genetic algorithms, and found them to be dominated by interface strategies tuned to payoffs. Critical realist strategies routinely go extinct. We do not need the stronger assumption that $X \subseteq W$, which is made by naive realist strategies, to have the interface strategies dominate.

Why, then, do we mention the stronger versions of veridical perception, given that they might lead to the confusion just discussed? Our goal is to provide a comprehensive formal description of the space of *all* perceptual strategies, so that none is inadvertently overlooked in current and future evolutionary analyses.

One might argue that, in our effort to be comprehensive, we spread the tent too widely, and included perceptual strategies that are mysterious. Cohen says in his note 5, "Additionally, if somehow, members of W were identical to members of X, that would make perception awfully mysterious: it's hard to imagine what causal mechanism could make it the case that distal world states (outside of heads) literally are states of cognitive systems (inside heads)." Perhaps it is hard to imagine. And we, like Cohen, are not fans of this perceptual strategy. But a failure to imagine causal mechanisms is insufficient grounds to reject a perceptual strategy from further consideration. To compare, people do not reject quantum theory, even though it is hard to imagine, despite decades of effort, what causal mechanism could explain entanglement or the outcome of a spin measurement. So we have opted to include in our classification of perceptual strategies even those strategies that we find implausible. Even the most implausible strategy can have its defenders. The claim that $W = X$ is, to us, implausible but not, perhaps, to an ontological solipsist.

In his note 4, Cohen mentions that we discuss notions of veridicality, such as homomorphic perceptual maps, that are weaker than the identity notion of veridicality that he (we think rightfully) impugns. But he then says, "Unfortunately, they don't say more about how they are understanding these weaker relations, or whether by invoking them they mean to be softening their pessimism about veridical representation."

But we do say precisely how we understand these weaker relations in our Definition 5 of critical realism. In this definition, we describe a weaker notion of veridical perception in which it is not the case that $X \subseteq W$—that is, in which no perceptual experience $x \in X$ is identical to any state $w \in W$ of the world, but in which the structural relations among the perceptual experiences $x \in X$ do mirror (i.e., are homomorphic to) the structural relations among the states $w \in W$. For instance, one might claim that color experiences x are not identical to any states w of the world, but that the relationships among color experiences nevertheless do faithfully mirror the true relationships among some states of the world. This would be an example of a critical realist version of veridical perception. Invoking this weaker notion of veridical representation does not at all soften our pessimism about veridical representation: These weaker forms of veridical representation also go extinct when we have them compete with interface representations that are tuned to fitness but that do not mirror the true relationships among states of the world.

(2) Cohen argues that we cannot claim that perceptions are falsidical without first specifying the content of perceptions. He says, "What is needed is the (standard) idea that perceptual states have content—intuitively, what they carry information about, tell us about, or say about, the world, and that can be evaluated for truth or falsity (the world either is as they say, in which case the states are veridical, or is not as they say, in which case the states are not veridical)." And he notes that it is an "apparent truism that you can't say whether something is veridical or not without first knowing what it is saying."

So Cohen proposes that we make a fatal mistake at the start of our project. We start by defining a perceptual strategy to be a map, $P : W \rightarrow X$, from states of the world W to perceptual states X. We never define the content of the states X—that is, what they are saying about W. Since we do not define what X says about W, we can't know whether X says something true or not. Thus we cannot distinguish veridical from falsidical perceptions, and a fortiori our evolutionary games cannot distinguish them, either. Our project is doomed from the start.

Fortunately, there is a remedy for this predicament: The well-established science of information theory can quantify precisely how much information a communication channel P can transmit from a source W to a receiver X. This theory *never* requires one to specify the content of the messages at the source or the receiver. Content is *irrelevant* to the quantification of information. One can assign *any* content that is compatible with the formal structures of W and X, without changing the channel capacity of P. The notion of information is a well-defined technical term that, along with other well-defined technical terms, such as entropy and mutual information and communication channel, allows one to formally analyze which statistical structures in the world are successfully communicated across a perceptual channel, and which are not.

This independence of content from the structure and quantification of information was a key insight that was critical to launching the modern field of information science. It is this insight that allows us to use the tools of information theory in our evolutionary games to reach sweeping insights into the evolution of perception, insights that do not require us to specify a priori the contents of the perceptual experiences X.

In particular, we can determine how much information a perceptual channel carries about both fitness and the state of the world, even if we never specify the content of X. If a perceptual channel carries *no* information about the state of the world, then a fortiori it carries *no veridical* information about the state of the world. Our evolutionary games indicate that natural selection pushes perceptual channels in the direction of carrying no information about the state of the world, and thus in the direction of carrying no veridical information about the world. Instead, it pushes perceptual channels toward carrying information about fitness that is just sufficient for survival. Content plays no role in the evolutionary games or in these conclusions; we informally ascribe content in particular examples to help the reader, but these are mere ladders to be kicked away once the example is understood. This is no surprise. Content plays no role in any of the theorems or technical advances of information theory. Similarly, in formal logic one need not know the contents of p and q to know that $p \vee q \Rightarrow p$ is a fallacy or that $p \wedge (-p \vee q)$ is false. "He who would distinguish the true from the false must have an adequate idea of what is true and false." Yes, but he does not need an adequate idea of what is the content of p and q.

Let's take this analysis one step deeper. Consider a world with states W and a generic fitness function f on W. Now, according to standard evolutionary theory, f depends on an organism, its state, and an action. But if we fix these, then we can view f as simply a function from W to, say, the real numbers $\mathfrak{R}$. If μ is a probability measure that describes the probability of states of W, then the probability of fitness states on $\mathfrak{R}$ is described by a *different* probability measure, usually denoted μf. Then it is straightforward to compute, for any perceptual channel P, the amount of information that P carries about μ versus how much it carries about μf. For a generically chosen f, the amounts will be different. This means that generically a perceptual channel will carry a different amount of information about fitness than it does about the state of the world. Our evolutionary games show that natural selection will push the perceptual channel to have more information about fitness and less about the state of the world. Again, the notion of information we use here is the technical notion used in information theory, which needs no claims about content.

(3) Cohen proposes content assignments for the perceptual states of dragonflies and jewel beetles so that their perceptions turn out to be veridical. In the case of the dragonfly, which sadly lays its eggs on oil slicks and gravestones that reflect horizontally polarized light, Cohen proposes that dragonfly perceptual states represent the content *horizontally polarized light at Location l*. In the case of the jewel beetle, which romances beer bottles that are dimpled, glossy, and brown, Cohen proposes that the beetle's perceptual states represent the content *displays a glossy, dimpled, brown surface*. With these content assignments, beetle and dragonfly both have veridical perceptions. Cohen then claims that we are not entitled to make content assignments so that their perceptions turn out to be nonveridical.

In fact, we do not need to. We do not need to make specific content assignments for which these perceptions turn out to be nonveridical. Our argument is different. We start with *no* content assignments to perceptual states. We start with a mathematical formulation of veridical and nonveridical perceptual strategies and their perceptual states (our Definitions 3–7) that requires no prior notion of content—in the same way that information theory needs no prior notion of the content of a set of messages before it can evaluate whether a communication channel veridically passes those messages. We then use the tools of information theory and evolutionary game theory to discover that the perceptual strategies, and therefore the perceptual states, that are favored by evolution are those that are *approximately* tuned to organism-dependent payoff functions—functions that *disappear* if the organism disappears.

Thus, Cohen's move will not work. Evolutionary theory entails that *all* properties of an organism's perceptual states are almost surely not veridical. This includes properties such as color, shape, glossiness, and location in space; it includes Cohen's *horizontally polarized light at Location l* and *displays a glossy, dimpled, brown surface*. Evolutionary theory entails, almost surely, that states of the objective world have no colors, shapes, surfaces, glossiness, dimples, locations in space, or horizontal polarization.

Here is one important consequence of this for philosophers: *Evolutionary theory constrains the admissible theories of perceptual content*. It rules out any theory of content in which properties of a perceptual experience—such as color, shape, illumination, motion, and their co-instantiation in objects—are taken to represent that *anything* other than the perceptual experience itself has such properties. Thus, it rules out contemporary theories of weak content, strong content, and rich content (e.g., Brogaard, 2014; Siegel, 2011). It rules out the direct realism of Searle (2015) and the disjunctivism of Campbell and Cassam (2014). Indeed, the challenge is to find any contemporary philosophical theory of the content of perceptual experiences that is not ruled out by evolutionary theory. The philosophical analysis of perceptual content must be restarted, using concepts and hypotheses constrained and motivated by the theory of evolution.

(4) Cohen argues that our evolutionary games show that perception veridically represents fitness payoffs, and therefore that we have not succeeded in demonstrating that perception is nonveridical.

We first correct a technical misunderstanding. After discussing our genetic algorithms for evolving perceptual strategies, Cohen claims, "it's not true that the evolutionary simulations under discussion compare realist, veridicality-requiring perceptual strategies against interface strategies that do not require veridicality. Rather, these games compare one veridicality-requiring strategy against a different veridicality-requiring strategy." Later, he adds that our "simulations only ever involve competition between perceptual strategies that represent veridically."

This is false. For the initial generation of the genetic algorithm, the set of strategies are chosen *randomly*, according to a uniform distribution. In all probability, *none* of the strategies in this initial generation are veridical, or even close to veridical. Instead, almost all strategies are unbelievably stupid and generate comically stupid foraging behaviors. It is only by breeding those that are slightly less stupid, and repeating this over hundreds of generations, that we find that selection favors strategies tuned to fitness. Strategies that veridically represent the true state of the world might never appear in any of the hundreds of generations. So, contrary to Cohen's claim that our simulations only ever involve competition between perceptual strategies that represent veridically, in fact in many of our simulations perceptual strategies that represent veridically never once appear.

Cohen then argues as follows:

> The alternative ("interface") strategy they pit against their realist strategy is no less committed to veridical representation. Of course, the states in the alternative strategy don't represent/carry information about/tell us about quantity of resources in each region, so they can't have resource quantities as their content. But these states do (by construction) represent/carry information about/tell us about the payoffs available in each region—which is just to say that those perceptual states have territorial payoffs as their contents. And those states are (again by HSP's own stipulation) veridical representers of the world: the payoff in a region just is exactly what the perceptual state represents it as being.

Again, we must make a technical correction. Evolution only produces *satisficing* solutions: An organism only needs to know a bit more than its competitors about fitness payoffs in order to outcompete them and drive them to extinction. The selection pressures are not to veridically perceive the fitness payoffs, just to perceive a tad more than the other organisms. Our evolutionary simulations show that perceiving a tad more about fitness is an advantage, but perceiving a tad more about the true state of the world is not.

But suppose that an organism does evolve to have perceptual states exceptionally matched to the fitness payoffs. Is it appropriate, then, to say, as Cohen does, "those states are . . . veridical representers of the world"? Not at all. Fitness payoffs are not states of the *world*, nor are they properties of states of the world. They are ephemeral *relationships* between an organism and states of the world. Remove the organism, and the fitness relationship disappears. Change the organism, and the fitness relationship changes. Perceptual states that represent fitness payoffs are not ipso facto representers of the world. Mathematically, a fitness function is a random variable whose distribution is, generically, utterly different than the distribution of states of the world. To veridically represent the fitness distribution is generically *not* to represent the distribution on the world. Cohen argues that "even if we take the simulations to show that evolution must have shaped perception to represent veridically the distribution of payoffs, they cannot show that it shaped perception to represent veridically only the distribution of payoffs." Yes they can. They can show that, *generically*, representing the distribution of payoffs precludes representing the distribution of states of the world. Here "generically" means except for fitness functions that are 1:1 with respect to world features, a vanishingly small subset of all possible fitness functions.

But still, one might reply, selection pressures are toward veridical representation of the fitness payoffs, even if these payoffs are not observer-independent aspects of the world. So isn't there still a sense in which perceptual states veridically represent, at least approximately, the fitness payoffs?

There is. But it is a sense quite alien to contemporary philosophical theories of perceptual content. In the weak, strong, and rich content theories, and in Searle's direct theory, a perceptual experience "as of" a dimpled, glossy, brown bottle shape a meter away is taken, in the normal case, to have the content that in fact something in the world is dimpled, glossy, brown, bottle-shaped, and a meter away, and would continue to be part of the world even if the observer and its experience disappeared. (Here the phrase "as of" is used by philosophers to refer to properties of the perceptual experience itself rather than to properties of an external or mind-independent object that is being perceived.) Presentationalist theories are similarly committed, in the normal case, to the persistence of something in the world that is dimpled, glossy, brown, and bottle-shaped in the absence of the observer.

The interface theory, to the contrary, claims that in the normal case a perceptual experience as of a dimpled, glossy, brown bottle shape a meter away does *not* have the content that a thing in the world is dimpled, glossy, brown, bottle-shaped, and a meter away, and would continue to be part of the world even if the observer and its experience disappeared. In fact this theory claims, on the basis of evolutionary theory, that predicates such as color, glossiness, shape, and distance in space are almost surely the wrong language to properly describe the world itself in the absence of the observer.

Such predicates can be useful representations of fitness payoffs, but again, in a sense alien to contemporary philosophical theories of perceptual content, such as the weak, strong, and rich content theories and Searle's direct theory. A perceptual experience as of a dimpled, glossy, brown bottle shape a meter away does not have the content that the fitness payoff is dimpled, glossy, brown, bottle-shaped, and a meter away. Fitness payoffs are abstract mathematical functions. Fitness is not colored, glossy, dimpled, or located in space-time; in this sense, our perceptual experiences are not veridical representations of fitness. However, our perceptual experiences as of colors, shapes, and locations are a satisficing graphical format, a serviceable user interface, that we can use to assess how to act in order to obtain enough fitness points.

The representational content of a blue rectangular icon on a desktop interface is not that in fact there is something blue and rectangular in the computer. It is rather that if you click *here* on the desktop, then you will open this file and see a photo, that if you drag *there* you will delete that same file, and so on.

Perceptual experiences are evolved conventions that allow us to get the fitness points we need to survive. Our mistake is to reify our species-specific conventions—to assume that a perception as of a dimpled, glossy, brown surface has the content, in the normal case, that a real surface in the world is dimpled, glossy, and brown. The contents of perceptual experiences are not about what exists in the world, they are about what to do to survive.

Shimon Edelman

Varieties of perceptual truth and their possible evolutionary roots. doi:10.3758/s13423-014-0741-z

Overview (1) Edelman states that we claim "the environment in which our species evolved is a highly structured place, containing many regularities." We reply that we made no such claim; that quote appears in our description of the standard Bayesian framework, which we describe simply in order to critique it and replace it. (2) Edelman defines the notions of categorical consistency, second-order isomorphism, and causality for perceptual channels. We show that categorical consistency corresponds to a special case of our notion of an interface strategy, and that second-order isomorphism and causality corresponds to a special case of our notion of a critical realist strategy. We provide a counterexample to his claim that categorical consistency entails veridicality. (3) Edelman states that the theory that underlies Roger Shepard's law of generalization both embraces evolution and predicts veridicality. We show that this claim is disproved by our evolutionary games and the invention-of-symmetry theorem. (4) Edelman states that the thesis of our article carries with it a lesson for studying the brain basis for vision and other faculties. We agree: The lesson is that the brain has no causal powers, and thus is not the causal basis for vision and other faculties. (5) Edelman asks, "Could true causal knowledge fail to be advantageous enough to become incorporated, under evolutionary pressure, into the cognitive toolbox of behaving animals?" We reply, "Yes."

We now consider these points in more detail.

(1) Edelman observes that we claim perception is nonveridical, that "perception is about having kids, not seeing truth." But then he says we try to eat our cake and leave it whole because he thinks we claim that "the environment in which our species evolved is a highly structured place, containing many regularities."

This is a simple misunderstanding. This statement about environmental regularities is not a description of our own theory, it is part of a section describing the standard Bayesian framework used by many perceptual scientists. We describe this framework in detail not to endorse it, but to critique it and then to provide a new framework that supersedes it. The interface theory of perception does not make specific claims about the existence of particular structures or regularities in the objective environment, other than the claim that this environment has the minimum probability-space structure required to discuss the probabilities of states.

(2) Edelman defines the notions of categorical consistency, second-order isomorphism, and causality for perceptual channels. He claims that these notions give us new ways that perception can evolve to be truthful, ways other than those that our article has exposed as absurd. His idea is that relationships among perceptions can veridically represent relationships in the objective world.

Unfortunately, Edelman's notions are special cases of the perceptual strategies we defined and tested in our article. We consider first his notions of categorical consistency (CC) and CC-truthfulness: "I call a perceptual channel . . . , denoted by a function $f : \mathcal{X} \rightarrow \mathcal{Y}$, CC-truthful if, when given as input any member x of a class of stimuli $\mathbf{X}$, $x \in \mathcal{X} \subset \mathbf{X}$, it reliably evokes a representation $f(x) = y \in \mathcal{Y}$. Here, $\mathcal{X}$ and $\mathcal{Y}$ are sets, each equipped with its own identity relation that defines set membership." This definition puts no constraints on the sets $\mathbf{X}$, $\mathcal{X}$, or $\mathcal{Y}$. If we assume, as is surely the case, that Edelman wants to be able to discuss probabilities, then these sets are measurable, and the definition of CC-truthfulness is a special case of our interface strategy.

A perceptual channel can satisfy Edelman's definition of CC-truthfulness and yet not allow veridical perception. An example is given in Fig. 3 of our article. Using Edelman's notation, we let $\mathbf{X} = \mathcal{X} = \{1, 2, \ldots, 12, 88, 89, \ldots, 100\}$, and $\mathcal{Y} = \{\mathbf{Red}\}$. Then, as is shown in Fig. 3, all elements $x \in \mathcal{X}$ map to the single element $y = \mathbf{Red} \in \mathcal{Y}$. We thus have well-defined set memberships for $\mathcal{X}$ and $\mathcal{Y}$, and the whole perceptual mapping satisfies Edelman's definition of CC-truthfulness. But one cannot tell from the percept **Red** the true state of affairs in the world: The world state could either be one of few resources (1, . . . , 12) or many resources (88, . . . , 100). So Edelman's definition of CC-truthfulness is simply a special case of perceptual strategies that we have already investigated in our evolutionary games. We have found that veridical versions of these strategies are not favored when they compete with nonveridical versions tuned to fitness.

Edelman's definitions of second-order isomorphism (SOI) and causality for perceptual channels are special cases of our critical realist strategies, in which metric distances and causal relations are preserved in the homomorphic mapping of the critical realist strategies. Once again, there are no new perceptual strategies here, just special cases of strategies that we have already tested and found unfavorable to veridicality. Edelman's hope that relationships among perceptions might veridically represent relationships in the objective world was tested extensively, and rejected, by our evolutionary games.

In the case of SOI, Edelman states that SOI-veridicality is generic if perceptual channels are smooth. He then observes that "the all-important questions here are, first, whether or not smooth perceptual channels become available often enough in the course of evolution, and, second, whether or not such channels survive competition with ones that are not smooth in the requisite manner. I conjecture that the answer to both questions is affirmative." We do not have to conjecture. We have used evolutionary game theory in hundreds of thousands of Monte Carlo simulations in which smooth channels compete with channels that are not smooth because they are tuned to fitness. Smooth channels generically go extinct. Therefore, SOI-veridicality also generically goes extinct.

(3) Edelman states that the theory that underlies Roger Shepard's law of generalization both embraces evolution and predicts veridicality. Shepard certainly did embrace evolution, asserting that our perceptual spaces have, in important respects, internalized the structure of the external world and its symmetries. But Shepard simply assumed that the external world is the space-time world of physics. He did not use the theory of evolution to show that this assumption is correct. When we actually test this assumption using evolutionary game theory and genetic algorithms, it proves to be generically false. Moreover, the invention-of-symmetry theorem shows that symmetries of perception of space-time, such as the symmetries that Shepard discusses, entail nothing about the structure of objective reality. Thus, although Shepard made (as it happens, false) assumptions about evolution, he in fact did not predict anything from the mathematical theory of evolution, and in particular did not predict veridicality from this theory.

Why is Shepard's law of generalization, and its later extensions by Tenenbaum and Griffiths (2001) and Chater and Vitányi (2003), so successful? Why are consequential regions connected, and why are generalization gradients smooth functions? The reason, contrary to the view of Shepard, is not that relationships among our perceptions preserve relationships in the objective world. Instead, the reason is that relationships among our perceptions *ignore* relationships in the objective world, and instead *preserve* properties of the relevant fitness functions.

For instance, looking at Fig. 3 of our study, one can see that each color in the perceptual space corresponds to a different category of expected fitness consequences. Red is one consequential region (bad consequences), yellow another consequential region (not quite as bad), blue another (good consequences), and green another (not quite so good). The consequential regions are connected not because the perceptual space preserves relationships in the objective world, but because it *wantonly rips up* those relationships in order to capture what is important: fitness.

So, for instance, the red consequential region rips up the relationships in the objective world by uniting resources of small quantity (0 through 12) with resources of high quantity (88 through 100). One cannot infer from the perception of red what quantity of resource is in the objective world, because the perceptual map destroys relationships that hold in the objective world in order to preserve fitness. It is because evolution tunes our perceptions to fitness that nearby perceptions have nearby consequences and that consequential regions are connected. In the perceptual space of this example, for instance, red is closer to yellow than it is to green because the fitness consequences of whatever is seen as red are closer to the fitness consequences of whatever is seen as yellow than they are to those of whatever is seen as green. This is the source of the connectedness of the consequential regions and the smoothness of the generalization gradient.

Shepard's law does not show that our perceptual systems have evolved to internalize external regularities of the objective world. To the contrary, it shows that our perceptual systems have evolved to ignore those regularities and to be tuned to fitness. Evolution has shaped perception so that nearby percepts have similar consequences—not so that nearby percepts have similar provenances in the objective world.

(4) Edelman states that the thesis of our article carries with it a lesson for studying the brain basis for vision and other faculties. We agree. The lesson is that our perceptual systems have evolved not to show us the true causal structure of reality, but rather to hide that causal structure, and to provide us with simplified guides to adaptive behavior. The appearance of causality, as when one object hits a second and appears to cause it to careen away, is just an appearance, a useful fiction. It allows us to predict future perceptions from current perceptions and actions. But in fact no objects in space-time have genuine causal powers.

And that includes the brain and its neurons. Neurons are simply a species-specific set of symbols that *Homo sapiens* happens to construct in dealing with an objective reality that is surely not spatiotemporal and that is surely utterly different and far more complex than neurons. For most normal neuroscience research, it is a harmless fiction to pretend that neurons have genuine causal powers. However, when we try to solve the mind–body problem—that is, to understand the relationship between neural activity and conscious experience—and then proceed to assume that neurons really do have causal powers and really do, somehow, cause our conscious experiences, then suddenly the fiction is no longer harmless. It has halted progress on the mind–body problem for decades, even centuries (Hoffman & Prakash, 2014).

So the thesis of our article does carry with it a lesson for studying the brain basis for vision and other faculties: There is no such brain basis. There is, of course, *some* causal basis, and *whatever* that basis really is, the best our perpetual systems can do when interacting with it is to come up with simple symbols that we call "neurons." We must not assume that the simple symbols our species has been shaped to employ in service of survival and reproduction are, in fact, insights into the true causal structure of reality. We perceive the world in terms of physical objects (such as neurons) residing in space and time not because this perception is a veridical reflection of true cause and effect, but because it happens to enhance our chance of successful reproduction.

(5) Edelman asks, "Could true causal knowledge fail to be advantageous enough to become incorporated, under evolutionary pressure, into the cognitive toolbox of behaving animals?"

We reply "Yes," and for the same reason that the windows interface on your laptop fails to reveal true causal knowledge about what happens inside the computer. You do not need to know all of the complex causality inside the computer to write a paper or edit a photo. Indeed, being forced to know the true causal structure of the computer would be an *impediment*. That is one of the main reasons we have an interface, so that we do not have to know and do not have to deal with all that causal structure.

Evolution has evolved perceptual systems that guide adaptive behavior. That *does* require that our perceptions allow us to *predict* the consequences of our behaviors. But this ability to predict is *not* the same thing as knowing the causal structure of objective reality. The ability to predict is simply the consequence of a good perceptual interface that can guide adaptive behavior while hiding all information, including causal information, that the organism does not need to explicitly know. Correlations among our perceptions, and correlations between our perceptions and the consequences of our behaviors, do not imply causation. Perceptual evolution is not about seeing true cause and effect, it is about having kids.

Jacob Feldman

Bayesian inference and "*truth*"*: a comment on Hoffman, Singh, and Prakash*. doi:10.3758/s13423-014-0795-y

Overview Feldman's basic position is substantially in agreement with ours. Like us, he believes that it is self-evident that fitness must necessarily dominate truth or veridicality whenever the two diverge. We focus our response on clarifying some finer points raised by his comments. (1) Feldman feels that our article gave somewhat short shrift to the question of whether true beliefs tend to facilitate fitness. We welcome the chance to further clarify our position on this important issue. (2) Feldman feels that Bayesian inference already embodies an epistemological stance similar to that of the interface theory of perception (ITP). We agree that a certain reading of some Bayesian authors can be interpreted in this way. We note, however, that most current Bayesian models of perception and cognition are routinely interpreted in a way that is far removed from the basic stance of ITP. (3) In passing, Feldman states that "a mental representation is a neural state, such as a particular pattern of neural activation." We note that, strictly speaking, ITP says that neurons are perceptual icons without any causal powers, and that therefore neural states cannot be identified with mental representations.

(1) Like us, Feldman considers it self-evident that fitness must dominate truth whenever the two are uncorrelated. He goes on to add, however: "The more difficult question is whether true beliefs tend to facilitate fitness. Hoffman et al. give somewhat short shrift to this question, setting up artificial games in which truth and fitness are decorrelated. The result—inevitably—is that fitness wins." If fitness does tend to be correlated with truth in natural environments, then it would seem that the evolutionary games whose results we cite—in which fitness functions are nonmonotonic, resulting in a "decorrelation" between fitness and truth—may seem somewhat artificial. Feldman goes on to argue that, in fact, he sees no reason to believe that the two are generally correlated, and he questions whether a statement expressing such a correlation is even meaningful. We agree, and elaborate on our own reasons—from the perspective of ITP—below.

There are two levels to consider. At one level, we can focus on a single relevant variable and ask whether an effective (i.e., fitness-enhancing) representation must preserve the underlying structure (say, the ordinal structure) of that variable. This is the level at which the evolutionary games we use posed this question. Their results clearly show that, whenever a fitness function on the relevant variable is nonmonotonic, an effective fitness-enhancing representation will generically not preserve the underlying structure of that variable. Within this context, Feldman's question may essentially be construed as: How common are nonmonotonic fitness functions? The short answer is: very. In general, we cannot expect payoff functions to vary monotonically with truth because, first, monotonic functions are an (unbiased) measure-zero subset of possible fitness functions and, second, independently of the first consideration, the ubiquitous biological need for homeostasis militates against monotonic fitness functions.

In addition to this specific level of analysis (focusing on the representation of a single variable), there is a deeper and more general reason why literally "true" beliefs do not tend to facilitate fitness. For this more high-level consideration, we return to the metaphor of the desktop interface. Two basic characteristics that make for an effective interface are abstraction and simplification. If a user interface were to literally depict the "true" state of the voltage in all circuits and transistors, it would actually be quite useless—a hindrance to the task of writing an article or editing a photo, rather than an aid. The usefulness of the interface lies precisely in the fact that it hides all of this "true" complexity of the computer from the user, and displays information only at a level of abstraction and simplification that is directly relevant to the goals of the user. The same is true, according to the ITP, of perceptual representations: They serve to provide an interface that supports effective interaction with the environment. The evolutionary pressures that shape the perceptual systems of a species have to do with its members' abilities to act in fitness-enhancing ways within their ecological niche. Faithfully representing the "true" environment—whatever that might mean exactly—is a goal quite far removed from these evolutionary pressures.

(2) Feldman feels that Bayesian inference already embodies an epistemological stance similar to that of ITP. We are inclined to say that Bayesian inference *potentially* embodies a similar stance. Some Bayesian authors of instrumentalist persuasion toward model selection (and the scientific enterprise in general) are indeed much more careful about distinguishing between models and "reality." All inferences for such authors are made on a space of models; "reality" or "the true state of the world" is kept out of the picture altogether. With such authors, including Feldman himself, we are very much in sympathetic accord. To such an understanding of Bayesian inference, what the ITP framework adds at a formal level is (a) the inclusion of fitness functions and perceptual channels as embodied in the computational evolutionary perception (CEP) framework (see Fig. 4b of the main article) and (b) the role of action in shaping perceptual channels (see the perception–decision–action [PDA] loop in Fig. 6 of the main article).

It is important to note, however, that when it comes to the modern literature on perception, Bayesian models are routinely used and interpreted in a way that is far removed from the basic stance of ITP. As we outline in the main article, current Bayesian models of visual perception standardly adopt an inverse-optics approach: The goal of visual processing is taken as "undoing" the effects of optical projection and "recovering" the true state of the world. Objects and their properties, such as shape and color, are taken not as the mind's model of the world, but as "true" intrinsic characteristics of the world itself. This essentially boils down to the naive-realist assumption that X is a subset of W (or that X is isomorphic to a subset of W). So, at the very least, it would seem that the calculus of Bayesian inference—which, as pure mathematics, is in itself perhaps neutral—can have a variety of epistemological and metaphysical assumptions attached to it. Our extension and generalization of the Bayesian framework, in terms of the CEP framework and the perception–decision–action loop, not only makes the epistemological stance explicit, but also provides a rigorous way to incorporate the role of fitness, and that of action, in the evolution of perception.

(3) In passing, Feldman states that "a mental representation is a neural state, such as a particular pattern of neural activation."

Although this was not a central point for Feldman, we would be remiss not to point out that one key implication of ITP is that, although it is a harmless fiction in everyday life to think and act as though physical objects have causal powers, nevertheless it is strictly speaking false to say that physical objects have causal powers, since physical objects are just a species-specific set of symbols that hide the truth. Thus, neurons, being physical objects, are without causal powers, although for everyday neuroscience research it is harmless, even useful, to pretend that they do. But when it comes to a fundamental understanding of perception and reality, and of the hard problem of consciousness, this fiction is no longer harmless, and in fact halts any further conceptual progress. So, strictly speaking, neural states cause nothing, and in particular do not cause, and are not identical with, mental states or mental representations.

Chris Fields

Reverse engineering the world: a commentary on Hoffman, Singh, and Prakash, "*The interface theory of perception*". doi:10.3758/s13423-014-0742-y

Overview (1) Fields suggests that ITP is supported by proven limits on solving the system identification problem. We agree, but consider a possible objection open to those who disagree. (2) Fields suggests that the notion of a hierarchy of virtual machines in computer science provides a useful way to understand ITP. We agree, but again consider a possible objection open to those who disagree. (3) Fields proposes that the perceptual interfaces of "higher" organisms can be expected to hide more information about the world than those of "lower" organisms. We reply that this is an intriguing proposal that deserves to be made mathematically precise and tested.

We now consider these points in more detail.

(1) Fields suggests that a result of relevance to ITP is a cybernetics theorem about the system identification problem—namely, the problem of discovering the proper description of a finite-state machine given only a finite set of observations of its behavior. The theorem states that this problem cannot, in general, be solved: The observations can set a lower bound on the complexity of the machine, but not an upper bound.

Fields points to the relevance of this theorem for ITP: "An organism cannot, therefore, determine by finite observations what the tokens comprising its own model of the environment refer to, and cannot determine what the tokens comprising any other organism's model of the environment refer to either. The best an organism can do is to construct an abstracted, meta-level model of its own or another's model of the environment." Krippendorff (2009) suggests that Ashby would agree: "Ashby always insisted that anything can afford multiple descriptions and what we know of a system always is an 'observer's digest.' "

This conclusion certainly comports well with ITP and suggests that it is worth exploring further how theoretical results in the field of cybernetics might constrain theories of perceptual evolution. In this regard, a couple of basic questions might provide a useful point of departure.

First, Krippendorff (2009) notes that "It is now recognised that the ability to determine the nature of a system by observation is limited to trivial machines." No one assumes that the objective world is a trivial machine. But one can ask whether it is plausible to think of the objective world as being usefully decomposable, at least in part, into trivial machines that are relevant to the fitness of an organism and whether, in such a case, evolution could shape the perceptions of that organism to determine the nature of those trivial machines. If so, the resulting perceptual strategy would not be a strict interface strategy, but a critical realist or even naive realist. This would not obviate the results of our evolutionary games that militate for ITP and against the evolution of critical realist and naive realist strategies. But it would entail that ITP does not accrue additional support from cybernetics. So it is a question worth addressing.

Second, Krippendorff (2009) notes that "Non-trivial machines, involving internal memories, defy observational determinability but can be understood by building them or taking them apart and reassembling them." And this gives another gambit for a critic of ITP. Organisms do not just observe the world, they interact with the world, sometimes by dissecting it. So (a) does cybernetics allow that organisms with the correct actions could eventually perceive objective reality as it is? And failing that, (b) does cybernetics allow that scientists, who routinely explore aspects of the world by "building them or taking them apart and reassembling them, " might at last understand objective reality? If point (a) is true, it would remove cybernetic support for ITP. If point (b) is true, it would not remove cybernetic support for ITP, but it would entail that ITP allows the possibility that science might eventually come to understand the nature of objective reality—a consequence we would in fact welcome.

(2) Fields suggests that the notion of a hierarchy of virtual machines in computer science provides a useful way to understand ITP. We agree. In fact, Fields's exposition on this point is a helpful unpacking of the interface metaphor of ITP.

It is hard to argue against the claim that virtual machines hide much of the reality of a computer from the user. However, one question a critic might raise is whether virtual machines really destroy all homomorphisms between the functioning of the computer and the functioning of the interface available to the user. If this critic should succeed, it would not count against the support of evolutionary games for ITP, but it would reveal a shortcoming of the interface *metaphor* in characterizing the kinds of perceptual strategies that evolution generically favors.

(3) Fields proposes that the perceptual interfaces of "higher" organisms can be expected to hide more information about the world than those of "lower" organisms. As Fields puts it: "we should expect 'higher' organisms, like 'high-level' programming languages, to encode less of the truth about the 'hardware' of the world, and to do so in a way that is more useful than the ways that 'lower' organisms do it."

This is an intriguing hypothesis, and deserves exploration. We would first want a quantitative measure of how much information a perceptual strategy has about the objective world—perhaps the mutual information between the probabilities of world states and the probabilities of perceptual states. We would then want a way to order perceptual strategies as "higher" or "lower," relative to another. For instance, we could say that a perceptual strategy P that can be written as a composition of perceptual strategies $P = P_1 \circ P_2 \circ \ldots \circ P_n$ is "higher" than each of the strategies in the composition: $P \geq P_i$, $i \in \{1, 2, \ldots, n\}$. Then we could try to prove, for example, that generically a "higher" strategy has less mutual information between perception and the world than its "lower" strategies.

However, it might turn out that careful analysis of perceptual evolution will reveal that, for generic fitness functions and generic probability measures on world states, the mutual information between perception and the world is 0, regardless of how "high" the strategy is. Clearly, these are interesting questions for further research, and attempting to answer them will deepen our understanding of perceptual evolution.

Jan Koenderink

Esse est Percipi & Verum est Factum. doi:10.3758/s13423-014-0754-7

Overview (1) Koenderink notes that there is a long history of the idea that perception is more like an interface than a window on reality. Koenderink sketches the history of this idea in the sciences, a history that we find to be stimulating and helpful. (2) Koenderink argues that the notion of a prior probability density on the physical world is useless, and even self-contradictory, and notes that we seem to endorse this notion. We clarify our view on this, which we believe aligns well with Koenderink's. (3) Koenderink argues that awareness, because it is not public, is outside the reach of science, whereas the physical world, because it is public, is within its reach. Here we might disagree, and we discuss our reasons.

We now consider these points in more detail.

(1) Koenderink notes that he is sympathetic with the interface paradigm for perception, and has used it himself in some of his publications. He points out that "the basic notions are hardly revolutionary. They occur in philosophy—that is protoscience—from the earliest days on. . . . As the sciences became established as such, the 'interface idea' kept surfacing." He then gives a brief history of the surfacing of the interface idea in the sciences.

We certainly agree that the interface idea is not new, and we appreciate Koenderink's brief history of the idea. Perhaps the key novel contributions of our article are to define precisely and comprehensively the possible perceptual strategies, and to demonstrate that natural selection, as modeled by evolutionary game theory, clearly favors strict interface strategies. In short, we give a rigorous evolutionary grounding for the interface idea.

(2) Koenderink argues that the notion of a prior probability density on the physical world is useless, and even self-contradictory, and notes that we nevertheless seem to endorse this notion. This is an important issue, and we welcome the chance to clarify our views.

First, we are not solipsists. We think that there is some kind of objective reality that goes beyond the perceptual experiences of a single observer. But the interface theory of perception is silent about the nature of that reality.

Second, we are Bayesians, in that we take probabilities to represent subjective degrees of belief. Accordingly, we do not propose that there is a true and objective prior probability of states of the world, but that attempts to think in terms of an a priori probability in the world will, generically, support interfaces and not veridicality.

We argue as follows. For *any* generically chosen probability of states of the world, and for any generically chosen fitness function, the theory of evolution by natural selection entails that a strict interface strategy will generically dominate, and will never be dominated by, any realist strategy whose perceptual space X has the same cardinality as the perceptual space of the strict interface strategy. This argument does not require commitment to any specific probability of states of the world, or to an objectivist interpretation of probabilities.

In fact, it does not even require that one postulate an a-priori-measurable structure on the states of the world. In our article, we define a perceptual strategy to be an onto measurable function $P: W \rightarrow X$, where $(W, \mathcal{W})$ denotes a measurable space of states of the world and $(X, \mathcal{X})$ denotes a measurable space of perceptual experiences. However, we could relax this definition and simply require that W be a set, and then induce a measurable structure onto W by pulling back the measurable sets of $(X, \mathcal{X})$ via P. Thus, we need postulate no a priori structure of any kind on W. (To say that themap P is *onto* means that its range is the *entire* set X, not some proper subset of X).

In short, we agree with Koenderink in rejecting the notion of an "objectively" existing prior probability density on the physical world.

(3) Koenderink argues that awareness, because it is not public, is outside the reach of science, whereas the physical world, because it is public, is within its reach.

Here we might disagree, because we do not grant that the physical world is public. Consider a pair of dice. You can roll the dice and we can both look and agree that they came up, say, with two 6s. This certainly suggests that the dice are public objects.

But they are not. Your perceptual experience that you describe as "dice that came up 6s" is numerically distinct from my perceptual experience that I describe as "dice that came up 6s." If this is not obvious, consider the Necker cube in Fig. 7 of our article. When you view this figure, you sometimes see a 3-D cube with the corner labeled "A" in front, but sometimes your perception flips and you see a 3-D cube with the corner labeled "B" in front. If you and a friend both look at Fig. 7, there will be times when you see "A" in front but your friend sees "B," clearly showing that your cube perceptions are numerically distinct. You see the cube that your perceptual system constructs, whereas your friend sees the cube that their perceptual system constructs.

The same is true of the dice. When you both look and agree that "the dice" have come up 6s, the dice of your perceptual experience are the dice that your perceptual system has constructed, and the dice of your friend's perceptual experience are the dice that their perceptual system has constructed. The experiences of the dice are numerically distinct.

Both you and your friend are interacting with an objective reality, in consequence of which you each construct the dice. But the theory of evolution entails that whatever that objective reality might be, it is almost surely not a pair of dice, nor anything like a pair of dice. The dice of your experience are simply the species-specific symbols you construct when interacting with that objective reality, whatever it might be. The reason that you and your friend can agree that the dice both came up 6 is that you are each members of the same species, with similar procedures for constructing your individual perceptual experiences. It is easy to move from "we both agree that the dice both came up 6" to "therefore we are both looking at exactly the same public dice." But that move is a logical error: Intersubjective agreement among perceptions does not entail the veridicality of those perceptions (because all of the perceptions could be nonveridical in the same way), and a fortiori does not entail the real existence of any public object answering to those perceptions.

In this sense, we say that there are no public physical objects. There is no public sun, moon, Mount Everest, New York City, electron, or Pacific Ocean. We each construct our own. We can talk about these objects and seem to communicate well, for the same reason that we can talk about our headaches. Our headaches are not public. I alone experience my headache, and you alone experience yours. But I assume that your experience is similar to mine so that we can talk meaningfully about our headaches. Similarly, I assume your experience as of the moon is similar to mine so that we can talk about "the moon" and agree, say, that tonight the moon is full. Our belief in public physical objects—that you see exactly the same moon as I—is a cognitive illusion based on a faulty inference.

Thus, science has never studied public physical objects because there are no such things. But science has progressed admirably nonetheless. Its success is not due to restricting its scope to public physical objects. Its success is due instead, at least in part, to the requirement of intersubjective agreement between experimenters. Here, such intersubjective agreement need not mean that two experimenters making measurements will have perceptual experiences with identical descriptions. For instance, two observers in different inertial frames measuring the length of a carrot will not arrive at the same answer. But their answers will be related by a Lorentz transformation and, in this more general sense, will still agree. They may also agree that they each see a carrot. But there is no public carrot, and no public length that they both measure.

As a result, we also disagree with Koenderink's claim that awareness, because it is not public, is forever beyond the scope of science. Science has done just fine with physical objects, despite the fact that they are not public, and there is no principled reason for science to be any less successful with awareness. In fact, we look forward to a vigorous science of awareness.

Perhaps we have misunderstood Koenderink on this point. If so, our apologies. However, the error we have just discussed is one we frequently encounter, so we can hope that this reply, even if it is off the mark for Koenderink's commentary, will be helpful to a broader audience. We consider this issue again in more detail in Point 2 of our response to McLaughlin and Green.

Rainer Mausfeld

Notions such as "truth" or "correspondence to the objective world" play no role in explanatory accounts of perception. doi:10.3758/s13423-014-0763-6

Overview (1) Mausfeld briefly reviews a long history of intellectual inquiry that concludes that perception is not veridical, or that the very notion of veridicality is irrelevant to understanding perception. We are grateful for this helpful review. (2) Mausfeld asks us to clarify our notions of the objective world, W, and perceptual states, X, to justify our use of measurable spaces to represent them, and to clarify what could be meant by the statements $X = W$ and $X \subset W$. We agree that these are critical issues, and we try to clarify them. (3) Mausfeld suggests that the vast majority of evolutionary change has little to do with natural selection, and that therefore our evolutionary games provide little insight into the real nature of perception and its evolution. We explain why we have studied the effects of natural selection on perception.

We now consider these points in more detail.

(1) Mausfeld says,

> Even a marginal consideration of what has been achieved during the long history of naturalistic enquiry into the nature of perception should make clear that issues of "veridicality" or of a "correspondence to the objective world" have, in the course of corresponding theoretical developments, been recognised as irrelevant for the explanatory purposes of perception theory. Hence, there is (or at least should be) no target left for the main thrust of HSP's arguments. But sadly enough, the actual state of affairs in perceptual psychology testifies to the contrary.

Mausfeld summarizes this enquiry, accenting his summary with several well-chosen quotes.

We find his summary helpful. We agree that a central claim of the ITP—namely, that our perceptions are not veridical—is not a new claim, but one that has been made repeatedly over many centuries.

(2) Mausfeld asks us to clarify our use of measurable spaces W and X to represent, respectively, the objective world and perceptual spaces, saying "the conceptualization of both W and X, and of their relation remains opaque to me." He then questions whether there can be any suitable notion of an objective world, noting that in philosophy, "we can find any number of metaphysical conceptions of what 'objective reality' is presumed to be. The same observation holds for the philosophy of physics." He also questions whether there can be any suitable notion of truthful perceptions. Finally, he says "Fortunately, these difficulties are of no relevance in the context of perception theory, where we aim at a theoretical understanding of the internal principles of a biological system."

These are important issues and we welcome the chance to clarify our views. We begin with Mausfeld's proposal that notions of objective reality and the truthfulness of perceptions are of no relevance to perception theory, and that the proper aim of perception theory is instead to understand the internal principles of a biological system. Space did not permit Mausfeld to spell out in his commentary what he means by the notion of "internal principles" that are the proper aim of perception theory. But his 2013 chapter tells us in some detail what he means, saying "perception can be understood as a triggering of conceptual forms by sensor inputs" mediated by internal evaluation functions that "integrate different internal sources of 'knowledge' about the potential causes for the activation of conceptual forms." He later elaborates, "The perceptual entities that make up our perceived world bear no relation of resemblance to the mind-independent entities by which they are elicited. They are rather, as particularly the Gestaltists have amply demonstrated, mental entities that are occasioned in the mind by suitable input conditions."

ITP agrees with Mausfeld that (generically) the perceptual entities of our phenomenal world do not resemble the mind-independent entities that elicit them. Moreover, ITP agrees with Mausfeld that a central task for perception researchers is to understand the eliciting process, not as an attempt to reconstruct mind-independent entities, but rather as a process endogenous to perception that follows its own internal principles. And ITP agrees with Mausfeld that, since perception is not about reconstructing mind-independent entities, there is an important sense in which the notions of an "objective world" and the "truthfulness of perception" are not central to perception theory. Thus it appears that the agreement between ITP and Mausfeld is substantial.

But it also appears that ITP and Mausfeld offer different, and mutually contradictory, reasons for concluding that the notions of an objective world and the truthfulness of perception are not central to perception theory. ITP offers the results of evolutionary games. Mausfeld argues instead that the question of whether perception mirrors the true structure of the objective world "will hardly survive the transition into a natural science context. It rather seems that no question remains that can be posed in a coherent and intelligible way. Hence the appropriate response to such a question is not to evaluate specific proposals but rather to dispel the delusion that an intelligible question has been raised."

In short, in response to the question "Is perception veridical," ITP answers "No" on the basis of evolutionary game theory, whereas Mausfeld asserts that the question is not intelligible. Why does he assert this? Mausfeld first notes that there is no consensus in philosophy as to what "objective reality" or "truth" is presumed to be. We agree with him on this point. He then says "Among the points of view developed in philosophy of science, the arguably most defensible forms of a scientific realism were instigated by ideas of Poincaré and Russell, and led to varieties of what is called 'structural realism,' " but "attempts at a coherent formulation are still encumbered with severe problems." Again, we agree.

But then Mausfeld concludes, "The theoretical clarifications achieved during the last centuries have clearly revealed that the notion of 'objective world' is of no relevance for the explanatory purposes of perception theory, and, in fact, borders on the inapprehensible, once we go beyond the context of ordinary discourse. This also applies to another fundamental notion of HSP's approach, namely 'truth,' whose meaning HSP likewise take to be self-evident." Here we disagree.

From the premises that (a) there is no consensus on the terms "objective reality" and "truth" and (b) theories of scientific realism have severe problems, the conclusion does not follow that the question "Is perception veridical?" is unintelligible and that the notion of "objective world" is of no relevance for the explanatory purposes of perception theory. A question can be intelligible even if currently has no satisfactory answer, and a notion can be intelligible even there is no agreement on its definition. This is standard practice in science. The notion of "gene" has been useful in biology for many decades, but its definition has repeatedly evolved and is still being revised (e.g., Gerstein et al., 2007). The notions of "space" and "time" have been useful in physics for centuries, but their definitions have also repeatedly evolved and, given that we lack a successful theory of quantum gravity, will surely undergo further revision. The notions of "objective reality" and "truth" and "veridical perception" have likewise been useful in perceptual science for centuries, but their definitions are still undergoing the healthy evolution that normally occurs in a vigorous science. Lack of consensus and an abundance of problems are not the hallmarks of irrelevance and unintelligibility, but rather of science at its rough-and-tumble best.

One source of our disagreement here might be two distinct notions of "explanation" that are relevant to perceptual theory: proximate and ultimate.

A proximate explanation describes *how* our perceptual systems *currently* function, including all of the internal principles, built-in concepts, and triggering mechanisms. This is clearly what Mausfeld has in mind when he says "in the context of perception theory . . . we aim at a theoretical understanding of the internal principles of a biological system."

An ultimate explanation describes *why* our perceptual systems function they way they do, and *why* they have *evolved* to employ particular internal principles, concepts, and triggering mechanisms. The search for ultimate explanations of perception is where the debate about veridicality arises: Did we evolve our internal principles, concepts, and triggering mechanisms as a result of evolutionary pressures toward veridical perceptions? The appropriate response to such an ultimate question is not to attempt to dispel the delusion that an intelligible question has been raised, but rather to recognize that this is a different sort of question than the proximal question, and that it is just as intelligible and amenable to scientific investigation as the proximal question.

Mausfeld himself seems to recognize the scientific intelligibility of questioning veridical perception when he says "The perceptual entities that make-up *[sic]* our perceived world bear no relation of resemblance to the mind-independent entities by which they are elicited" (Mausfeld, 2013. The legitimate ultimate question is: For what evolutionary reason did the internal principles and triggering mechanisms of our perceptual systems evolve so that the perceptual entities that make up our perceived world bear no relation of resemblance to the mind-independent entities by which they are elicited? We propose in our article that a profitable approach to studying this ultimate question is to use the tools of evolutionary game theory, evolutionary graph theory, and genetic algorithms.

Mausfeld questions the way that we formulate the perceptual strategies that compete in our evolutionary games and genetic algorithms. We define a (dispersion-free) perceptual strategy, *P*, as a surjective (i.e., onto) measurable function $P : W \rightarrow X$, where $(W, \mathcal{W})$ denotes a measurable space of states of the world and $(X, \mathcal{X})$ denotes a measurable space of perceptual experiences. Mausfeld says of this,

> HSP base their analysis on a notion of "objective world," whose meaning they apparently presume to be self-evident and a matter of course. They also take this meaning to be precise enough to be mathematically idealisable in terms of a set *W* and a σ-algebra **W** over *W*, yielding measurable events (which they need in order to get the probability spaces underlying their dynamic models). Accordingly, they formalise the "objective world" by a measurable space (*W*, **W**)."

Actually, our motivation is different. We do not take the notion of "objective world" to be self-evident and precise. Instead, we assume that the notion of "objective world," like the notions of "space" and "time" and "gene," has some provisional meaning that is to be refined as science advances. In this context, we propose to model the objective world as a measurable space because this is the *least restrictive* mathematical formulation that allows perceptual *events* to be somehow related to objective world *events*. If we do not have a measurable structure, then there are no objective events and no perceptual events, and therefore no hope of experimental tests of theories of the relationship between perceptual events and events in objective reality. So our approach is to propose the least restrictive formalization of objective reality that still allows empirical science to be possible. This makes our mathematical model of "objective reality" as provisional as possible, and avoids accidentally ruling out a priori certain mathematically precise meanings for "objective reality" that the advances of science might discover to be most useful. It is for such reasons that we do not stipulate other structures on W, such as a topology, a metric, a partial order, or even a specific probability measure.

The happy outcome is that even with this most general and provisional mathematical formulation of the objective world and perceptual experiences, it is still possible to rigorously define the notion of a perceptual strategy and to exhaustively classify a nested hierarchy of such strategies. This allows us to address the evolutionary question of perceptual evolution in a mathematically precise manner, where the mathematical formulation has not inadvertently precluded any potentially relevant possibilities from the outset. If we had defined the objective world to be, say, the 11-dimensional space-time of certain quantum string theories, we would have ruled out a priori many interesting and potentially relevant possibilities.

Mausfeld then says "Furthermore, from the given definitions of W and X, I find it incomprehensible how a relation between the two can be of the type 'correspondence', 'being caused by', or 'representing'. This applies, again, in particular for the cases $X = W$, or $X \subset W$."

We included these cases to be comprehensive. We do not want to preclude any possibilities from consideration, even possibilities that we might find intuitively implausible.

The case $X = W$ would presumably be endorsed by a metaphysical solipsist, who claims that no reality exists other than his or her own mind or mental states. Should this case be confirmed by experiment, we would thereby learn that "objective reality" W consists entirely of mind or mental states, and that it is not distinct from perceptual states X. If we had a priori required them to be distinct, or had required W to be mind-independent, we would have precluded such a discovery. The relationship between X and W would not be one of "correspondence," "being caused by," or "representing"; it would be *identity*. And that is why we must not, a priori, specify what the relationship between X and W must be. Instead, we must let our best science tell us, eventually, how best to construe that relationship.

The case $X \subset W$ would presumably be endorsed by a direct realist such as Searle (2015). Searle prefers not to call the relationship between X and W a representation, but rather a *presentation*. Should Searle's position be confirmed by experiment, we would learn that the relationship between X and W is not one of "representing," but rather of "presenting."

ITP says that, generically, $X \not\subset W$, and that the only structures preserved by the perceptual mapping are the measurable structures. In this case, we can think of perceptions being triggered by objective reality in much the same way that Mausfeld does, and we can take perceptions to be representational. But what they represent would be aspects of the relevant fitness functions rather than aspects of objective reality.

Hopefully these examples give the flavor of our mathematical approach. We want rigor: Our mathematical definition of the various perceptual strategies gives that. We also want the least restrictive mathematics, so that no potentially relevant perceptual strategy, no matter how counterintuitive, is a priori precluded from analysis: Using only measurable spaces and maps gives us that. We leave the nature of objective reality as an open question to be settled, not a priori by mathematical assumption, but a posteriori by science. Thus, some of the questions that Mausfeld wants answered now are left purposefully unanswered at this point, so that they can be answered authoritatively by the eventual progress of our best science.

(3) Mausfeld suggests that the vast majority of evolutionary change has little to do with natural selection, and that therefore our evolutionary games provide little insight into the real nature of perception and its evolution. He says

> Another crucial notion in HSP's analyses is that of natural selection, which they apparently and without further discussion regard as an almost exclusive factor regulating evolutionary change. In fact, however, the actual role of natural selection in the evolution of complex biological systems is far from obvious. . . . Evolutionary biology has, in more recent years, accumulated pervasive evidence that suggests that the vast majority of evolutionary change has rather little to do with natural selection.

Indeed, the debate between "adaptationists" and "pluralists" goes back at least to Ronald Fisher (1930) and Sewall Wright (1931). Adaptationists place greater emphasis on the role of natural selection in evolution, whereas pluralists emphasize the role of other factors in evolution, such as genetic drift, constraints from physics and biochemistry, and developmental and genetic constraints (such as linkage and pleiotropy) on genetic variation and transmission (see, e.g., Garvey, 2007; Godfrey-Smith, 2014; Orzack & Sober, 2001).

We are not extreme adaptationists. We recognize that the relative influences of natural selection, genetic drift, physics, biochemistry, development, linkage, pleiotropy, and other factors must be analyzed case by case. Our research so far has focused on the influence of natural selection on perceptual evolution, in part because the standard evolutionary arguments for veridical perception have been based on natural selection—namely, that veridical perceptions are fitter. We have not encountered arguments that veridical perceptions are favored by genetic drift, constraints from physics and biochemistry, or developmental and genetic constraints on genetic variation and transmission. However, it will be fascinating to explore how all of these factors and their interactions influence the evolution of perception.

Mausfeld says "I know of no arguments from evolutionary biology that support HSP's claim that 'the distinction between fitness and truth is central to evolutionary theory'. Rather, 'truth' is a notion that, for principled reasons, cannot be expected to play a role in explanatory frameworks of evolutionary biology." We are surprised by this comment. It is standard in evolutionary biology to claim that a physical object, say a piece of chocolate, is truly part of objective reality, but that this object can have different fitness consequences for different organisms—for instance, chocolate is poisonous to cats but safely consumed by humans. Here there is a clear distinction between what the evolutionary biologist takes to be the "truth"—that is, to be a genuine part of objective reality—and what the evolutionary biologist takes to be the fitness consequences of that "truth." Perhaps Mausfeld is simply claiming that evolutionary biologists do not typically talk about "truth." That's fine. We are simply using the term "truth" here as a shorthand for "true state of the objective world," a notion that evolutionary biologists take for granted when they assume that organisms, genes, and food sources really do exist in the objective world.

Brian P. McLaughlin and E. J. Green
Are icons sense data? doi:10.3758/s13423-014-0780-5

Overview (1) McLaughlin and Green (hereafter, MG) question whether we can use evolutionary theory to support our claim that space-time and physical objects are the wrong language for describing the objective world. They note that evolutionary theory assumes, contrary to our claim, that organisms and other physical objects exist in space-time. We reply by reviewing the logical structure of our argument, and showing that it is not self-refuting. (2) MG suggest that we miss a basic distinction between an experience and what it is an experience of when we say, "When you and I both look at your car, the car I experience is not numerically identical to the car you experience." They worry that the ITP thereby risks endorsing idealism. We reply by clarifying our ideas with an example from virtual reality, and showing how ITP is not committed to an idealist, or any other, ontology. (3) MG suggest that ITP needs to embrace the sense-datum theory of perception as the only way to properly understand its metaphor of desktop icons. We reply by expanding on our reasons for not embracing the sense-datum theory, and for keeping the icon metaphor.

We now consider these points in more detail.

(1) MG raise an important point about the logical structure of our argument. We use the theory of evolution to argue for the conclusion that space-time and physical objects are the wrong language for describing the objective world. But the theory of evolution itself assumes that there are physical objects in space-time, such as organisms and food. Moreover, the notion of a fitness function, which figures centrally in our arguments, depends critically on the organism. So it certainly looks like we are using the theory of evolution to refute the theory of evolution, and thus logically shooting ourselves in the foot.

But we are not. The key is that the theory of evolution has a logical core, sometimes called "universal Darwinism," which captures the essence of evolutionary theory—namely, replication, retention, and selection. This logical core is elaborated in the mathematics of evolutionary game theory, evolutionary graph theory, and genetic algorithms. Universal Darwinism is domain neutral, and even ontologically neutral in the following sense. It describes the evolution of competing *strategies*, but is neutral about what *substrates* instantiate those strategies. This is one of Dennett's (1995) key points in *Darwin's Dangerous Idea*, and a key reason he describes universal Darwinism to be a "universal acid" that transforms an ever-widening range of domains, allowing the development of new fields such as memetics.

So, although the classical theory of biological evolution is committed to a physicalist ontology, and to the real existence of physical organisms, universal Darwinism is not so committed. It is ontologically neutral in such a way that permits the physicalist ontology of the classical theory to be plunged into a bath of its universal acid, to discover just how much of that ontology survives. The answer is clear: Almost surely nothing of that ontology survives; it all dissolves away. Almost surely the language of our perceptions, the language of space-time and physical objects, is a species-specific adaptation that simply guides adaptive behavior. No selection pressures shape the language of our perceptions to be an appropriate language to describe objective reality. Universal Darwinism tells us that the notions of organism, species, resources, and so on must ultimately be understood in a different ontology for which space-time and physical objects are not an appropriate language. But universal Darwinism does not tell us what that ontology may be.

So our argument is not self-refuting. The notions of organism, reproduction, and heredity in universal Darwinism are not committed to any ontology; they are not committed to physical organisms and physical notions of heredity and reproduction. We can, without fear of logical self-contradiction, pour the ontologically neutral acid of universal Darwinism over the ontology of space-time and physical objects, and find that this ontology disappears.

This highlights the power of evolutionary theory. Suppose it *could not* be used to properly answer the following question: Has the language of our perceptions—the language of space-time and physical objects—been shaped to describe objective reality? This is a meaningful and important question about the evolution of our perceptions. So if the theory of evolution could not answer it, or could in principle only answer "Yes," then we would surely demand a real theory that could give us a genuine answer. Fortunately, universal Darwinism is a real theory and gives a genuine answer. However, it is a shock to our intuitions that the answer is "No."

But perhaps not a shock in retrospect. Perhaps we shouldn't be surprised that a species would mistake its perceptions for objective reality. Indeed, perhaps most species never recognize that there might be a distinction between perception and reality, and simply take their perceptions to be identical to objective reality (or simply have no conception of a reality beyond their perceptions). Perhaps there are no selection pressures to make the distinction; the ability of our species to make this distinction, and then to discover the shocking news, is simply a spandrel, an unselected ability that is an accidental consequence of some other cognitive capacity that was selected.

(2) In our article, we claim "When you and I both look at your car, the car I experience is not numerically identical to the car you experience." MB respond

> Why is the car that you visually experience, that you see, not numerically the same car that I visually experience, that I see? Can't we see exactly the same car, for instance, my car sitting in the driveway? (I might want to buy your car. I certainly don't want to buy my perception of your car.) We think that HSP are missing a familiar basic distinction here. There is a distinction between an experience (or perception) and what it is an experience (or perception) of.

However, it is precisely this distinction, combined with the results of our evolutionary games, that leads us to our claim. An example from virtual reality (VR) should help to clarify our claim. Suppose you and a friend decide to play tennis in VR. You each put on a helmet and body suit, and find yourselves immersed in a VR tennis game on a hard court. You hold a green felt tennis ball, which you serve and your opponent smashes back for a winner.

First, we claim that your perceptual experience as of a green felt tennis ball is numerically distinct from the perceptual experience that your friend has as of a green felt tennis ball. We do not take MB to disagree with this claim. One bit of evidence for the claim is that it is possible for you to close your eyes and for your perceptual experience as of a green felt tennis ball to cease, whereas at the same time your friend's eyes remain open and his perceptual experience continues. So your experiences are numerically distinct.

Second, and more controversially, we claim that there is no objective, physical, green felt tennis ball that you and your friend both see. You both talk, of course, about "the tennis ball" that you served and he smashed back for a winner, as though you both see a single tennis ball in objective reality. You know that such talk is not literally true, but that it is a harmless fiction for the purposes of playing the game, and has the advantage of economy of expression. There is an objective reality with which you and your friend are interacting, and this objective reality makes possible the coordination of perceptual experiences that you and your friend have as of a green tennis ball, so that you can enjoy hitting "the ball" back and forth. But, in this VR example, this objective reality is not a green tennis ball—it is transistors, voltages, magnetic fields, and software.

So it is possible in VR to hit tennis balls, shoot snakes with guns, race cars, and drive red fire engines, even though the objective reality that makes this possible features none of these items. ITP says the same is true in everyday life.

This VR example still does not address another part of MB's objection, because in the VR example, the objective reality—namely, a high-powered VR computer—is still a reality embedded in space and time; some aspect of this space-time reality, say the states of registers in the VR computer, is probably homomorphic to the perceived experiences of tennis balls, snakes, guns, and so on. So the VR example does nothing to help suspend disbelief in our claim that the language of space-time is almost surely not the right language for describing objective reality. As MB say "Moreover, space-time exists and there are physical objects in space-time. Were there no such thing as space-time, then general relativity theory would be false, or at least inapplicable to 'the objective world'. "

So we need an example of an objective reality that is not space-time but that in principle could do everything that space-time does, including coordinating the perceptual experiences of different observers so that they can enjoy the useful fiction that they are interacting with a public physical object, such as a tennis ball or a snake. Fortunately, physicists in search of a theory of quantum gravity have done the hard work for us here. To take just one example, Seth Lloyd (2005) has shown that one can start with *abstract* quantum bits and quantum gates—*not* in any way located in space and time—and from them construct the space-time of general relativity. As Lloyd (2005) puts it, "The basic idea of the 'computational universe' research program proposed here is that what happens to quantum information is fundamental: all other aspects of the universe, including the metric structure of spacetime and the behavior of quantum fields, are derived from the underlying quantum computation. To paraphrase Wheeler, 'it from qubit'. " Abstract qubits are proposed to be the fundamental reality, and all of space-time and the "its"—that is, the physical objects—within that space-time are constructed upon this qubit foundation. Space-time and physical objects, in Lloyd's theory, are *not the right language* for describing objective reality. This precisely accords with the prediction of our evolutionary games.

So, take the case of a particular "it" such as, say, a snake. Lloyd's theory of quantum gravity provides a precise theory of objective reality in which it is possible for two observers to agree that they both see a snake, and in which it is possible for them to enjoy the useful fiction that they both see the same snake and shoot the same snake, and yet the objective reality that makes this possible cannot be described in the language of space-time and physical objects. Rather, the language of space-time and physical objects is derivative upon a more fundamental objective reality that is properly described with an entirely different language—namely, the language of quantum information.

We are not, of course, claiming that Lloyd is right. That remains to be seen. But his theory of quantum gravity shows how, in principle, objective reality might not be describable in the language of space-time and physical objects, but might require a new language that describes in a more fundamental way the "its" and laws that science has so far tried to describe using the old language of space-time and physical objects.

MB state,

> It should be noted that it would be a mistake for a quantum theorist to deny the reality of middle-sized objects. As John Bell pointed out, there must be be-ables. Quantum mechanics is the most highly confirmed physical theory we've ever had. But it is not possible to describe the evidence we have for it without saying things that entail the existence of various kinds of physical objects. We can't justifiably appeal to evidence obtained from particle accelerators, for instance, without presupposing that there are particle accelerators. If there are no particle accelerators, then, trivially, there is no evidence obtainable from particle accelerators. We can't justifiably appeal to evidence involving light beams passing through slits in walls, if there are no walls; and so on.

This argument misunderstands how science really advances. Yes, in the process of constructing a new theory, such as quantum theory, physics looks to evidence from, for example, particle accelerators and walls with slits. But one consequence of the new theory might be that we must completely reconceptualize the nature of particle accelerators and walls with slits. Our old conceptualizations motivated our experiments, and were the essential ladder to the new theory. But once we have the new theory, that theory itself might instruct us to kick away the ladder, or our conceptualization of that ladder. That is part of the powerful cognitive leverage that science confers. It actually has the potential to correct deeply mistaken conceptualizations of the very data that inspired its theories. We saw this in the case of evolutionary theory. It was originally formulated in a framework that took for granted the objective existence of organisms, replicating molecules, resources, and so on. But the theory of universal Darwinism that arose out of these conceptualizations in fact has the power to turn around and apply its acid to these conceptualizations, and to completely dissolve them away. Science, in short, really has the potential to teach us something new, even something new about the very way we conceptualize the ontology of our data and experimental apparatus. A theory of quantum gravity based on quantum computation might justify one to reject the reality of middle-sized objects, even if the assumption of the reality of middle-sized objects was an essential conceptual step in the ladder that led to the theory of quantum gravity. No logical contradiction is thereby entailed.

(3) MG suggest that ITP should embrace the sense datum theory of perception. They argue that, in so doing, ITP would reap important benefits:

> By endorsing sense datum theory, HSP can avail themselves of the sense data theorists' explanation of how we can have knowledge of the objective world, even though we are only ever directly aware of our own, private sense data. By endorsing the theory, they can as well avail themselves of the techniques sense datum theorists developed for showing how the fact that we are directly aware only of our sense data coheres with our ordinary, everyday views about there being trees, rattlesnakes, deserts, cameras, driveways, cars, and the like.

We certainly appreciate the helpful spirit in which this suggestion is offered. But as we have just discussed above in Point 2 of our reply to MB, we doubt that our perceptual experiences, in the normal case, give us any knowledge of the objective world. Our evolutionary games and genetic algorithms make it clear that our perceptual experiences instead give us satisficing information about the fitness consequences of our actions, information that is quite different than knowledge of the objective world. In this regard, ITP does not cohere with "our ordinary, everyday views about there being trees, rattlesnakes, deserts, cameras, driveways, cars, and the like." Instead, ITP corrects this ordinary everyday view, much like Eratosthenes' computation of the circumference of the earth corrected the ordinary, everyday view that the earth is flat, and the observations of Galileo corrected the ordinary, everyday view that the earth is the unmoving center of the universe. Sometimes it is best not to cohere with ordinary, everyday views. For most everyday purposes it is of course true that ordinary, everyday views such as a flat earth, a geocentric universe, and objective rattlesnakes are harmless, even useful, fictions. But for more demanding purposes, such as circumnavigating the earth, sending a rocket to Mars, or understanding perceptual evolution, these fictions are no longer harmless. They obstruct conceptual and technological advances.

Of the four reasons we give in our article for denying that ITP is a sense datum theory, only our fourth reason—our claim that undergoing a phenomenal state does not involve a two-place, act–object relation—is one that MB agree would militates against ITP being a sense datum theory. So we will focus on this reason. MB argue that we are wrong to claim that ITP does not posit a two-place, act–object relation in perceptual experience:

> When they speak of seeing a rattlesnake, they suggest that the rattlesnake that we see is an icon. *Prima facie*,

> these descriptions purport to denote objects to which we stand in perceptual or experiential relations. Moreover, consider the following. Either we are aware of icons or we are not. If we are not aware of icons during perceptual experience, then in what sense do we encounter them? What role do they play in perceptual experience? If, however, we are aware of icons, then we are either directly aware of them or only indirectly aware of them (that is, aware of them only by means of being aware of something that is not an icon). If we are only indirectly aware of icons, then what are we directly aware of? If we are aware of anything, there must be something that we are directly aware of. If we are directly aware of icons, and icons are, as HSP seem to hold, mental entities, then icons are sense data.

The claim of this kind of argument, as Searle (2015) shows, turns on an ambiguity. As he puts it: "this claim is ambiguous because it contains two senses of 'aware of,' which I will call, respectively the 'aware of' of intentionality and the 'aware of' of constitution." Searle goes on to note, "I am aware of a visual experience, but this is a totally different sense from the intentionalistic sense because, to repeat, the visual experience is identical with the awareness itself; it is not a separate object of awareness."

Searle's point is that if we are aware of a visual experience as of a rattlesnake, then there is no need to posit a sense datum object. The visual experience is *identical* with the awareness itself, and is not a separate object of awareness. If I have a visual experience as of a car, or a gun or rattlesnake, this does not entail that there must be car, gun, or rattlesnake sense data that are the objects of the experience. These visual experiences as of cars and guns are what we refer to as "perceptual icons."

We hasten to note, however, that although we agree with Searle's analysis here, and with his dismissal of sense data, we do not agree with his further conclusion that perception, in the normal case, directly presents reality as it is. His theory of direct realism in perception is flatly contradicted by the theory of evolution.

MB ask, "what *are* the members of the set X of possible experiences? Our contention is that HSP's remarks on this issue seem to commit them to the view that the members of X are episodes of being aware of perceptual icons, and that perceptual icons are mental entities. This looks to be sense datum theory."

Not at all. Each $x \in X$ is this: a perceptual experience as of x. For instance, consider a theory of the perception of 3-D objects in motion based on the rigidity theorem of Shimon Ullman (1979). Ullman's theorem requires three orthographic images of four noncoplanar points to infer whether a perception as of a rigid 3-D motion is possible, and if so, what that 3-D perception is (along with its orthographic reflection). In the simplest case, the space of possible perceptions X is thus a 3 (frames) × 4 (points) × 3 (dimensions) = 36-dimensional real Euclidean space. Thus, each $x \in X$ is mathematically a point in $\mathfrak{R}^{36}$. It stands for a perceptual experience as of four points in three-space undergoing a certain motion over three discrete instances of time. Nearby points in X are similar perceptual experiences as of points in motion. No sense data are implied or required.

MB then ask, "First, where are perceptual icons located? Are they located in the brain, or, failing this, within some other structure in objective reality? If so, then why do they not count as parts of objective reality? If not, are they instead in a type of private mental space? If so, what are private mental spaces?"

The answer depends on what we mean by "objectivity." As Tyler Burge (2010) points out, this word has many different meanings. For instance, if we are talking about perception representing the world objectively, then "objectively" means, roughly, accurately or veridically. But if we are talking about the objectivity of subjective matters, then "objectivity" might mean (a) mind independence, (b) lack of representational content, or (c) all that is real. If we are talking about the objectivity of empirical measurements, then "objectivity" might mean intersubjective agreement.

If MB are asking whether perceptual icons are part of objective reality in the sense of "all that is real," then we would answer "Yes," that we take perceptual experiences, and therefore perceptual icons, as part of objective reality. In particular, we take perceptual experiences as of brains—that is, brain icons—as part of objective reality. But if the question is, Does a brain *icon* represent the world objectively—that is, veridically—then we would answer that the theory of evolution dictates that the answer is almost surely "No," and thus, furthermore, that brains (as opposed to brain icons) are almost surely not part of objective reality in the sense of "all that is real." Thus, perceptual icons are almost surely not located in brains.

Does this entail that perceptual icons are located in a private mental space? ITP itself does not answer that question, because ITP is not committed to any particular ontology (e.g., physicalist, dualist, or idealist). If one augments ITP with, say, Seth Lloyd's ontology of qubits and gates, and then adopts, say, Tononi's (Oizumi, Albantakis, & Tononi, 2014) integrated information theory of consciousness, in which the amount and kind of conscious experience depends on the amount and kind of integrated information in a system, then one might get an answer about where perceptual icons are located that physicalists might find congenial, even though the language of spacetime and physical objects has been replaced by the language of qubits and quantum gates. We are not endorsing this ontology, but simply pointing to it as a possibility. We happen to be pursuing a different ontology (e.g., Hoffman & Prakash, 2014). But the key insight of ITP—that our perceptions are almost surely tuned to fitness rather than to objective reality—can be cashed out with many different theories of what that objective reality might be.

MB ask:

> Second, if perceptual icons are indeed located in a private mental space, can they interact with constituents of objective reality? There is reason to think they must. HSP's PDA loop posits a mapping (D) from perceptions (X) to actions (G), and a further mapping (A) from actions to states of the world (W). This indicates that they allow that our actions can exert influence on the objective world, and that our actions are guided by our perceptions (and hence by our awareness of perceptual icons). Thus, perceptual icons must be capable of interacting with objective reality. But how can elements of a private mental space interact with things from outside that space?

Indeed, whether or not perceptual icons end up being located in a private mental space, the PDA loop does assume that perceptual experiences guide actions, and that actions can influence the objective world. Just how this works in detail will depend on what ontology one adds to ITP. In an ontology generated by qubits, quantum gates, and information integration, the actions and influences would all be consequences of, or ways of describing, quantum information processes.

MB ask:

> Third, assuming that perceptual icons in fact interact with constituents of objective reality, will a completed physics need to take icons into account? If so, why don't such icons count as parts of objective reality after all? If not, then must we admit that physics is not a causally complete theory, that there are factors that cause motion that physics does not take into account?

We have already discussed the first part of this question. However the question of causality deserves further comment. Causality is a concept critical to science but still lacking both a precise definition and a consensus on its metaphysical status (e.g., Beebee, Hitchcock, & Menzies, 2009). Some contend that causality refers to a transfer of energy, or counterfactual dependence, or statistical covariation, or the instantiation of laws or regularities, or a fiction of the human mind, or some combination of these. ITP does not entail commitment to any specific theory of causality. Its requirement that perceptual maps be measurable functions guarantees that perceptual events are functionally related to events in the objective world (technically, that the pull-back of each perceptual event is a world event). But it guarantees no specific pattern of covariation between probabilities of world events and probabilities of perceptual events, in part because perceptual maps are influenced by fitness functions that can radically alter such covariation.

However, ITP accepts the verdict of evolution that there are no selection pressures for veridical perceptions, and that therefore our perceptual experiences as of physical objects in space-time are almost surely not veridical: Nothing in objective reality has the property of being a physical object in space-time. This entails that nothing in objective reality has the property of being a physical object in space-time that has causal powers. Note that this conclusion does not require commitment to a particular theory of causation.

So we accept that physics is almost surely not causally complete. This conclusion comports well with some interpretations of quantum theory (e.g., Fuchs, 2010), and with theories such as Seth Lloyd's (2005), which posits that fundamental causal relations reside entirely in a nonspatiotemporal world of quantum logic gates.

However, ITP does allow that perceptual *experiences* as of physical objects in space-time can have causal powers. Indeed, it assumes that evolution has shaped such perceptual experiences to *guide* adaptive behaviors. The precise concept of cause that underlies such guidance is, we readily admit, an important and open issue for further research.

It is perhaps not such a big leap from the views of MB to those of ITP. MB say that they "think there is positive reason to believe that normal perception is typically nonveridical in certain respects", and they allow for the possibility that

> Nothing in the objective world has the property of redness. Sense datum theorists allow, however, that things in the objective world can count as red in a derivative sense, namely in virtue of being disposed to cause us, in appropriate circumstances, to be aware of a red sense datum. Thus, they say that it is true that the fire engine is red. It's true that the fire engine is red, because the fire engine is disposed to cause a normal perceiver to be aware of a red sense datum under normal circumstances.

Issues about sense data aside, ITP agrees that nothing in the objective world has the property of redness, and that the objective world can dispose us, in appropriate circumstances, to have perceptual experiences as of red. ITP simply takes this a step further: Nothing in the objective world has the property of being a fire engine, but the objective world can dispose us, in appropriate circumstances, to have perceptual experiences as of a fire engine. Nothing in the objective world has the property of being space-time, but the objective world can dispose us, in appropriate circumstances, to have perceptual experiences as of space-time. What MB can accept for color, ITP accepts as well for objects and space-time, because the theory of evolution gives strong reason to treat all of them, equally, as species-specific guides to adaptive behavior.

Zygmunt Pizlo

Philosophizing cannot substitute for experimentation: comment on Hoffman, Singh & Prakash. doi:10.3758/s13423-014-0760-9

Overview (1) Pizlo argues that there is no logical connection between 3-D vision and other senses. In particular, he feels that outside of 3-D vision, there is not much of a role for a priori constraints, since such constraints are usually simply not available. We disagree. One of the main overarching themes to emerge in the perceptual sciences over the last few decades is that perception, in all its forms, involves inductive inference based on a priori constraints (although the strength of the available prior constraints may well vary across perceptual systems and modules). (2) Pizlo claims that veridicality is a purely empirical question, and that theoretical analysis ("philosophizing") has no role to play. We disagree. Words like *veridical* may seem perfectly innocuous and well defined, but they actually hide ambiguities and unquestioned assumptions. It is precisely the goal of theoretical analysis to bring these to the fore. Our analysis clarifies, for example, that what is generally termed "veridicality" is really a form of coherence between different forms of measurement—and *not* any sort of match with the "true" state of the world. (3) Pizlo thinks that what our ITP offers is a regression to the motor theory of perception (MTP). This is simply a misunderstanding of ITP. What ITP claims is that action plays an important role in the *evolution* of perception (i.e., in the shaping of perceptual channels over the course of evolution). ITP does *not* claim that action is required in order to perceive, as MTP does. Despite surface-level similarities (mainly because of the involvement of action), ITP has in fact very little in common with MTP. (4) Pizlo says: "it is the shape of 3D objects that conveys the information we all use to perceive our world veridically. . . . Specifically, 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." It is certainly intuitively appealing to think that, because our perceptual experiences contain specific regularities and invariances, the objective world W itself must contain those regularities. But our *invention-of-symmetry theorem* shows such arguments to be entirely baseless—proving that almost nothing can be inferred about the structure and regularities in W on the basis of the invariance properties of our perceptions and actions.

We now consider these points in more detail.

(1) Pizlo argues that there is no logical connection between 3-D vision and other senses. In particular, he feels that outside of 3-D vision, there is not much of a role for a priori constraints, because such constraints are usually simply not available. We disagree. One of the main overarching themes to emerge from the perceptual sciences in the last few decades is that perception, in all its forms and varieties, can be fruitfully viewed as a form of inductive inference. The form of inference is inductive because—no matter what the perceptual modality or module—the *inputs* to perceptual systems do not logically determine a unique *output* or *interpretation*. Perceptual processing can resolve this ambiguity only by bringing to bear, implicitly or explicitly, additional prior constraints. These constraints are derived from prior experience—whether phylogenetic or ontogenetic—with the environment, and are embodied in the way in which sensory information is processed. Such considerations have led to the rise in prominence of Bayesian approaches to provide a unifying scientific framework for perceptual science.

It does not follow from this, of course, that all prior constraints are equally strong—that is, equally good at resolving the ambiguities inherent in perceptual problems. Indeed, the calculus of Bayesian inference captures precisely the extent to which a prior constraint, or a combination of prior constraints (suitably expressed as probability distributions), will resolve a perceptual ambiguity. This alone is sufficient to explain why powerful constraints such as symmetry can allow the visual system to infer invariant 3-D interpretations from 2-D images (as has been shown by Pizlo and colleagues' elegant empirical research on this topic). But such claims are quite distinct from a claim of veridicality—because the 3-D interpretations are part of the perceptual interface, not of objective reality. Similarly, the a priori constraints used to resolve inductive ambiguities are not, in ITP, regularities internalized from the 3-D physical world—again because the "3-D world" is part of our perceptual interface, and not of objective reality.

(2) Pizlo claims that veridicality is a purely empirical question, and that theoretical analysis has no role to play. We disagree. Words such as "veridical" may seem perfectly innocuous and well defined, but actually they hide ambiguities and unquestioned assumptions. It is precisely the goal of theoretical analysis to bring these to the fore. "Veridical" means "corresponding to the true state of affairs," which, in the context of perception, boils down to "corresponding to the true state of the mind-independent world." However, if we look at how perceptual psychophysicists often substantiate claims of "veridicality" (of 3-D shape, say), we find that what they in fact do is examine the degree of coherence between two different forms of measurement: (a) psychophysical measurements of an observer's percept of 3-D shape, and (b) spatial measurements taken in the physical (or a simulated) environment. But the latter measure is decidedly *not* part of the objective, mind-independent world W; it is still part of the species-specific perceptual interface of *Homo sapiens*. To be more precise, it corresponds to what we call the "measured world," M—a cognitive representation obtained as a result of applying scientific measurement procedures. Although the measured world M clearly goes beyond our perceptual representations, it is really simply an *extension* of our perceptual representations. For, on the one hand, scientific procedures clearly allow us to arrive at measurements that could not have been arrived at

from raw perception alone; and yet, the *formal structure* within which these measurements are placed is determined by the way in which *we* as *Homo sapiens* perceive the world (e.g., as having a 3-D structure). These 3-D perceptions are then extended to a cognitive model of space, often by making certain symmetry assumptions. For example, if we assume that our 3-D framework must be invariant under translations and rotations, then our perceptual representations of space would get extended to a Euclidean framework (as in Newtonian physics).

Thus, when we compare psychophysical measurements of shape to spatial measurements in the physical (or a simulated) environment, we are simply evaluating the degree of coherence between two different levels of description within our perceptual interface. This can indeed be an informative evaluation. But we are not somehow getting outside of our own interface in order to compare perceptual experience with objective reality.

(3) Pizlo thinks that what our ITP offers is a regression to the MTP. He feels that, because our framework involves a role for action, it automatically falls under a class of theories for which action is required in order to perceive—that is, for which there can be no perceptual experience without concomitant action ("sitting is seeing," as he puts it). This is simply a misunderstanding of ITP. In ITP, action plays an important role in the *evolution* of perception—that is, in the shaping of perceptual channels over the course of evolution (and, more generally, in the shaping of PDA loops). This is quite distinct from any claim that action is required in order to perceive.

To see this explicitly, note that in the PDA formalism, the *existence* at time t of an organism's perceptual experience $x_t \in X$ is not *contingent* on the decision D or action A of that organism, nor is x_t formally *identical* to D or A. Of course, the action A of an organism can alter the state of the world W, which in turn can alter the perceptual experience x_t via the kernel P. But this is just to say that action can *influence* perception, not that action *is* perception. A theory that precluded such influence would be prima facie false.

This misunderstanding leads Pizlo to, rather strangely, compare ITP with Watson's extreme behavioristic position of "thought as implicit speech" and the dark ages of radical behaviorism. Nothing could be further from the truth. Contra behaviorism, ITP involves a fundamental role for *representation* and *probabilistic computation*. Moreover, ITP provides a precise mathematical framework in which fitness features in a central way, thereby allowing one to capture the *evolution* of perception. These characteristics place ITP squarely within the realm of modern computational theories of mind—and about as far away from behavioristic approaches as is possible. Indeed, we have previously referred to this approach as *computational evolutionary perception* (see Hoffman & Singh, 2012; Singh & Hoffman, 2013).

(4) Pizlo says: "It is the shape of 3D objects that conveys the information we all use to perceive our world veridically. . . . Specifically, 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."

Let us take these claims at face value—that is, as being claims about objective reality. These claims then essentially boil down to the idea that, because our perceptual experiences contain specific regularities and invariances, we can therefore infer that the objective world W itself must contain those regularities. Such arguments certainly have an intuitive appeal, but our article shows them to be entirely baseless. Indeed, our invention-of-symmetry theorem shoots an arrow through the heart of such arguments—showing that almost nothing can be inferred about the structure and regularities in W on the basis of the invariance properties of our perceptions and actions. Given how directly relevant this theorem is to Pizlo's claims, it is surprising that his commentary makes no attempt to grapple with it.

Matthew Schlesinger

The interface theory of perception leaves me hungry for more. doi:10.3758/s13423-014-0776-1

Overview (1) Schlesinger notes that we postulate a PDA loop operating at both phylogenetic and ontogenetic time scales, and he suggests that we should explore how the interaction of these time scales affects evolution. (2) He proposes that the question remains open whether the "perception" of lawful perception–action couplings (a.k.a. "affordances") is in fact directly experienced. (3) Schlesinger suggests that our evolutionary games and genetic algorithms could benefit from having more realistic simulations of embodied organism with active sensorimotor systems. (4) He suggests that it would be helpful to have more anecdotes about real animals whose perception is determined by fitness rather than by objective reality, anecdotes that would illustrate how this can be an advantage. (5) Schlesinger notes that we say nothing about qualia, and suggests that since ITP already presupposes that perception and action are dynamically linked, it is only one small step forward to suggest that perceiving these lawful (or contingent, or statistically predictable) regularities is ipso facto what constitutes the qualities of perceptual experience. (6) Schlesinger suggests that our evolutionary games need to go beyond studying the evolution of phenotypes in isolation, and to address the full nonlinear complexity of the evolution of interacting structures and behaviors. We agree substantially with Schlesinger on each point, and will try to clarify ITP by discussing each one.

We now consider these points in more detail.

(1) Schlesinger notes that we postulate a PDA loop operating at both phylogenetic and ontogenetic time scales, and he suggests that we should explore how the interaction of these time scales affects evolution—for example, through the Baldwin effect. We completely agree.

One step in this direction was taken by Brian Marion (2013) in his Ph.D. dissertation at UC Irvine, with experiments on chromatic adaptation. As Marion states in his abstract to chapter three,

> Humans rapidly adapt to the local background chromaticity, leading to improved color discrimination near the adapted chromaticity, at the cost of reduced discrimination sensitivity in other parts of the visible spectrum. This process has generally been understood to be automatic and fully task-dependent. In this study, we present the first evidence that chromatic adaptation can be manipulated by placing a utility structure on the subject's performance. When participants received a greater reward for successfully discriminating among shades of blue than among shades of red, their performance on blue trials improved significantly more than their performance on red trials, despite otherwise identical exposure and task characteristics. These findings suggest that the process of chromatic adaptation works to actively optimize expected gain, rather than as a simple automatic response to chromatic stimuli. (p. 53)

Marion's result does suggest how the PDA loop operating at phylogenetic and ontogenetic time scales could profoundly affect the path of evolution. His results suggest that the sensitivity of human perceptual discriminations can be retuned in just a few minutes in response to changes in environmental fitness functions. This ontogenetic ability to retune would confer a fitness advantage in the context of environments in which fitness functions change appreciably in space or time. In such environments, one would expect that this ability to quickly adjust to changing fitness functions would improve reproductive success, and thus would affect the genetic makeup of the species. That is, one would expect the Baldwin effect.

This does raise a question for ITP. Is it possible that selection pressures on a perceptual system to be ontogenetically responsive to a rapidly changing environment are ipso facto selective pressures toward veridical representation of that environment? Perhaps the nonveridical interface strategies of ITP are only favored in slowly changing environments? If the jewel beetle, for instance, could have dealt more flexibly with a changing environment, perhaps by having more veridical perceptions, it might not have faced extinction at the hand of a beer bottle.

This certainly deserves careful study. One would want to distinguish situations in which the environment changes rapidly and fitness functions do not from the converse situation in which fitness functions change rapidly and the environment does not. And, of course, one would also want to study the case in which both change rapidly. We predict that in the generic case in which fitness functions do not vary monotonically with structures in the objective environment, nonveridical perceptions tuned to fitness will dominate veridical perceptions.

(2) Schlesinger proposes that the question remains open whether the "perception" of lawful perception–action couplings (a.k.a. "affordances") is in fact directly experienced. With some terminological reservations, we agree. Our evolutionary games show that perceptions are almost surely not direct experiences of any objective structures of the world. They also show that selection pressures push our perceptual experiences to be *satisficing* representations of the relevant fitness functions. So, if our perceptions are direct experiences of anything, they might be direct experiences of fitness consequences in the exercise of the PDA loop.

The catch is the qualifier *satisficing*. This is not merely an obligatory academic hedge; it has real bite. It is quite possible that our perception in terms of, for example, shapes, colors, and motions is not a perception tuned to any one specific fitness function, but rather is a satisficing representational system that is adequate for a wide variety of relevant fitness functions, but optimal for none. In this case, our perceptual experiences might be direct experiences of fitness consequences, but not of the consequences tied to any one particular fitness function.

Our terminological reservation is about the word "affordances," which has been given a specific meaning within the naive realist framework of Gibson, a framework that ITP does not share. Reading "affordances" as "fitness consequences" should avoid confusion here.

(3) Schlesinger suggests that our evolutionary games and genetic algorithms could benefit from having more realistic simulations of embodied organisms with active sensorimotor systems. For instance, he suggests having at least a very basic sensorimotor system—for example, an eyeball with a 1-D or 2-D retina, situated in a simple environment, and that can change its visual input by rotating.

Again, we agree. A first step in this direction has been taken by Justin Mark (2013) in his Ph.D. dissertation at UC Irvine, with a genetic algorithm that evolved perception and action strategies in a foraging game. In some of the simulations, organisms were only allowed to "look" in some directions and not others, crudely simulating an eyeball with directional gaze. Mark found that this did not change his fundamental result—namely, that the perceptions of the later generations in his genetic algorithm evolved to be tuned to fitness, and that veridical perceptions simply did not evolve at all. But this is a first step, and more work clearly needs to be done using more comprehensive simulations of embodiment and, indeed, robotic systems with genuine embodiment.

(4) Schlesinger suggests that it would be helpful to have more anecdotes about real animals whose perception is determined by fitness rather than by objective reality, anecdotes that would illustrate how this can be an advantage. As he says,

"The goal, of course, of these colorful scenarios would be to provide an intuition pump that mapped our own everyday experiences into the world of these imaginary creatures, and ultimately, gave us a handle on what it would mean that our perception is determined by fitness rather than by objective reality."

Once again, we agree. Our examples of the jewel beetle and dragonfly were intended to be such intuition pumps, but more examples would be helpful.

However, two points are worth highlighting here. First, if ITP is correct, then every perception of every creature is, almost surely, determined by fitness rather than by objective reality. So one might think, given ITP, that we could pick any perception at random and use it as an intuition pump.

But, of course we can't. Indeed, to the contrary, most examples of perception are taken as prima facie evidence for the opposite intuition—namely, the intuition that perception is determined by, and accurately represents, objective reality.

This raises an obvious problem for ITP: Why should good intuition pumps be so rare? And this leads to the second point. What are we actually doing when we use an intuition pump such as a jewel beetle humping a bottle? Are we comparing the beetle's perceptions with objective reality and showing that they fall short? Well, if ITP is correct, then we almost certainly are not doing that, because we almost certainly have no idea what the objective reality is. The objective reality is not a bottle. We laugh at the jewel beetle, and say that it mistakes a bottle for a female. But that is not correct. It is more correct to say that the male beetle mistakenly takes something to be a female beetle that *H. sapiens* takes to be a beer bottle. The point is that, according to ITP, *H. sapiens* no more sees objective reality here than does the jewel beetle. The perceptions of *H. sapiens*, like those of the jewel beetle, are only satisficing representations of species-specific fitness contingencies. It just so happens that, in this particular case, the perceptions of *H. sapiens*—but not those of the male beetle—indicate that no fitness points can be reaped by trying to mate.

And that is the key to a successful intuition pump. The key is not to find situations in which we see truly and some poor besotted creature does not; we see no more truly than other creatures. Instead, the key is to find a situation in which our perceptions happen to discriminate fitness consequences that the perceptions of another creature, to its detriment, do not. This rarely happens—which is why intuition pumps are hard to come by—and when, as usual, it does not happen, we naturally assume that we, and in its own way the creature, both see objective reality. But this assumption is just as insightful as the beetle humping the bottle.

(5) Schlesinger notes that we say nothing about qualia, and suggests that since ITP already presupposes that perception and action are dynamically linked, it would only be one small step forward to suggest that perceiving these lawful (or contingent, or statistically predictable) regularities is ipso facto what constitutes the qualities of perceptual experience.

Schlesinger is right that we say nothing about qualia in our article. Partly that is because qualia are sometimes taken to entail ineffability and sense data, so we have preferred to speak of *perceptual experiences* rather than qualia. In our description of the PDA loop, the elements $x \in X$ denote perceptual experiences.

We agree that it is an intriguing idea that the qualities of perceptual experience might be constituted, at least in part, by the lawful regularities in the dynamical interplay of perception and action. This is surely worth pursuing. The bigger project here is to understand the relationship between phenomenology and perceptual content, which is a big open question in the philosophy of perception (e.g., Brogaard, 2014; Campbell & Cassam, 2014; Searle, 2015; Siegel, 2011). Much of the current work on this problem is influenced by the assumption that perception is, in the normal case, veridical. It is likely that ITP, which denies this assumption and asserts instead that perception is tuned toward fitness, will lead to a novel account of the relationship between phenomenology and perceptual content.

(6) Schlesinger suggests that our evolutionary games need to go beyond studying the evolution of phenotypes in isolation, and to address the full nonlinear complexity of the evolution of interacting structures and behaviors.

We absolutely agree. We expect that we will continue to find that veridical perceptions go extinct when they compete with nonveridical perceptions tuned to fitness. But that is just the first baby step toward a theory of perception informed by the theory of evolution. The full richness of the competition and evolution of perceptual interfaces has yet to be explored. Having taken the first baby step, we can now begin to develop a genuine theory of perceptual evolution.

Author note This research was supported in part by grants from the National Science Foundation and the Federico and Elvia Faggin Foundation.

References

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

Beebee, H., Hitchcock, C., & Menzies, P. (Eds.). (2009). *The Oxford handbook of causation*. Oxford, UK: Oxford University Press.

Brogaard, B. (2014). *Does perception have content?* Oxford, UK: Oxford University Press.

Burge, T. (2010). *Origins of objectivity.* Oxford, UK: Oxford University Press.

Cabello, A., Estebaranz, J. M., & García-Alcaine, G. (1996). Bell–Kochen–Specker theorem: A proof with 18 vectors. *Physics Letters A, 212,* 183. doi:10.1016/0375-9601(96)00134-X

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

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences, 7,* 19–22.

Dennett, D. (1995). *Darwin's dangerous idea*. New York, NY: Simon & Schuster.

Fisher, R. (1930). *The genetical theory of natural selection*. Oxford, UK: Oxford University Press: Clarendon Press.

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

Garvey, B. (2007). *Philosophy of biology*. New York, NY: Routledge.

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., . . . Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Research, 17,* 669–681.

Giustina, M., Mech, A., Ramelow, S., Wittmann, B., Kofler, J., Beyer, J., . . . Zeilinger, A. (2013). Bell violation using entangled photons without the fair-sampling assumption. *Nature, 497,* 227–230. doi:10.1038/nature12012

Godfrey-Smith, P. (2014). *Philosophy of biology*. Princeton, NJ: Princeton University Press.

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

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

Krippendorff, K. (2009). Ross Ashby's information theory: A bit of history, some solutions to problems, and what we face today. *International Journal of General Systems, 38,* 189–212. doi:10.1080/03081070802621846

Lloyd, S. (2005). The computational universe: Quantum gravity from quantum computation. arXiv:quant-ph/0601135.

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

Mark, J. (2013). *Evolutionary pressures on veridical perception: When does natural selection favor truth?* Irvine: Ph.D. Dissertation. Department of Cognitive Sciences, University of California.

Mausfeld, R. (2013). The attribute of realness and the internal organization of perceptual reality. In L. Albertazzi (Ed.), *Handbook of experimental phenomenology: Visual perception of shape, space and appearance* (pp. 91–118). Chichester, UK: Wiley.

Mermin, N. D. (1985). Is the moon there when nobody looks? Reality and the quantum theory. *Physics Today, 38*(4), 38–47. doi:10.1063/1.880968

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology, 10,* e1003588. doi:10.1371/journal.pcbi.1003588

Orzack, S. H., & Sober, E. (2001). *Adaptationism and optimality*. Cambridge, UK: Cambridge 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,* 515–519.

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,* 791–794.

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,* 220404.

Searle, J. (2015). *Seeing things as they are: A theory of perception*. Oxford, UK: Oxford University Press.

Siegel, S. (2011). *The contents of visual experience*. Oxford, UK: Oxford University Press.

Singh, M., & Hoffman, D. D. (2013). Natural selection and shape perception: Shape as an effective code for fitness. In S. Dickinson & Z. Pizlo (Eds.), *Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective* (pp. 171–185). New York: Springer.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*, 629–640.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

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.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics, 16,* 97–159.