Imitation and Extinction: The Case Against Reality

This piece appears in the Los Angeles Review of Books Quarterly Journal: Imitation, No. 23

To receive the Quarterly Journal, become a member or purchase at our bookstore.

¤

Imitation is the cleverest form of weaponry deployed by predators and prey. Mexican milk snakes, which lack venom, sport vivid bands of red, black, and yellow that imitate the deadly Texas coral snake — a lie that shields their helpless young from raptors. The Australian katydid, *Chlorobalius leucoviridis*, imitates the alluring clicks of female cicadas, and feasts on males it seduces. In such mimicry, natural selection conscripts imitation in service of deception.

It's natural to suppose, conversely, that selection also enlists imitation in service of truth — creating a systematic match between our perceptual experiences and the real world. I see a green pear. Does the shape and color that I experience match the true shape and color of the real pear? If I'm sober, and don't suspect a prank, then yes, of course. I assume that my pear experience is "veridical": it imitates properties of the real pear. If I close my

eyes, my experience morphs to a gray field, but the pear, I presume, is still there.

Most experts agree. The neuroscientist David Marr argued that perception "delivers a true description of what is there." Robert Trivers, an evolutionary theorist, concurs that "our sense organs have evolved to give us a marvelously detailed and accurate view of the outside world." The cognitive scientist Anthony Chemero adds that "the objects of our accurate perceptions … exist in an animal-independent world." That is, the objects of our accurate perceptions are real, in the sense that they exist even if no creature perceives them.

¤

Why should evolution favor perceptions that are veridical? Stephen Palmer, a cognitive scientist, explains that, "[e]volutionarily speaking, visual perception is useful only if it is reasonably accurate." The idea is that our predecessors who saw more accurately outcompeted rivals who saw less accurately, thus raising their chance to become our ancestors (rather than just dust) and bequeath to us their genes, which coded for more accurate perceptions. Countless generations of such competition are congealed in our genes, ensuring that our perceptions imitate reality — not all of reality, just what we need to survive.

This logic seems compelling, but it's flawed. Natural selection now has a precise formulation known as evolutionary game theory. Applying game theory to evolution means we can run simulations and prove theorems about its effects. The results are clear. Veridical perception does not assist with survival. In fact, an organism whose perceptions are veridical is less fit than its nonveridical twin whose perceptions are focused on what we call "fitness payoffs," a central concept in evolution that we'll explore next. In short, seeing the truth will make you extinct.

Why? Natural selection is about scoring fitness payoffs. Think of video games in which players hunt for points, trying to reach the next level. In evolution, creatures hunt for fitness payoffs. If you wrangle more payoffs than the competition, you're more likely to launch offspring into the gene pool of the next generation.

In video games, the points you can grab depend on your tools. With a magic key, you might raid a treasure chest; with a protective shield, you might brave a flamethrower. In evolution, the fitness payoffs depend on the organism and its state. For a hungry person looking to eat, eucalyptus leaves are deadly. For a hungry koala looking to eat, those leaves offer high payoffs. For a sated koala looking to mate, they offer none. For us, air with too little or too much oxygen is fatal; only a narrow range of oxygen is just right — a Goldilocks effect common to many fitness payoffs. For a trout, sustained immersion in air is always fatal. Such examples illustrate that fitness payoffs are not indelible inscriptions in the granite of reality, but ephemeral relations between a creature and its environs. Billions of species have graced the earth; 99 percent of them, and their payoffs, have gone extinct.

Natural selection has shaped our senses to present fitness payoffs. Only if these payoffs, in turn, track the world (as, for instance, the height of a column of mercury tracks temperature in a thermometer or tracks air pressure in a barometer) will our senses track the world, and thereby be veridical. How likely is that? Research shows the probability is zero: to track the world, payoffs must satisfy exacting equations. So, senses that hunt reality are like chess players that hunt pawns: they're playing the wrong game.

Our senses do not present "a marvelously detailed and accurate view of the outside world" as Robert Trivers writes. They present payoffs. The very language of our senses — space, time, objects, shapes, colors, tastes, smells, and sounds — is simply the wrong language to describe reality; its

vocabulary was not shaped to that end. (What is the right language? We don't know, but later I'll sketch an idea.)

¤

This conclusion is stunning, and raises a natural question: how can our senses be useful if they aren't veridical?

A metaphor might aid our intuitions. Suppose you're playing *Grand Theft Auto*, and your car has a black steering wheel. As you sit in your ride and peer out the windshield, the wheel is in the lower left corner of your screen. Does this mean that a black wheel lurks inside the lower left corner of your computer (or Playstation or Xbox)? Of course not. The video interface does not show the truth — which, in this metaphor, is circuits, volts, and software. In fact, the interface expressly hides this complexity; instead, it shows icons that let you control circuits without even knowing they exist. If you had to flip bits to turn the steering wheel, you wouldn't drive the car.

That's what evolution did: it endowed us with perceptions that are an interface, not a true depiction of an animal-independent world. Spacetime is our 3-D desktop and physical objects are 3-D icons. Our perceptions of spacetime and objects hide reality, whatever it is, and provide tools to interact with reality despite our ignorance of its nature. What we normally take to be reality is, in fact, a simplified virtual reality, shaped by natural selection to guide adaptive action; it keeps us alive long enough to raise offspring. We think that perception imitates an animal-independent world. Instead it creates a virtual reality whose details depend crucially on the needs of the animal. This is the key idea of the "interface theory of perception," or ITP.

Why, then, can we agree about what we see? If I see a red grape on a table, chances are so can you. How so, if there is no real grape, and we just see icons in interfaces? Isn't it simpler to admit that there is a real grape that we

both see? According to ITP, we agree because we are conspecifics — we belong to the same species — with interfaces that are similar in format and engagement with reality, whatever reality might be. In multiplayer *Grand Theft Auto*, Joe in New York and Janet in Montreal can agree that Joe is leaning on a red roadster because they share a similar interface and interact with the same hidden realm of circuits and software. There's no need to posit a real roadster that both players see. Consensus is just consensus; it's not proof of what reality actually is. Agreement isn't insight.

ITP claims that cars are just icons. If so, then why can they hit us and kill us? Surely, they're more than icons.

Evolution shaped us with icons to keep us alive. We must take them seriously or die with no offspring. But this gives no reason to take them literally. Suppose I've written a book, and the icon for its file is red, rectangular, and in the center of my desktop. If I carelessly drag that icon to the trashcan, I could lose the book and months of work. I should take the icon seriously. But the book is not literally red, rectangular, or in the center of my computer.

Doesn't evolution assume that physical objects, such as DNA, are real? If we use evolution to prove that objects are just icons, haven't we used evolution to disprove evolution — a logical contradiction?

Not at all. Evolution by natural selection is now a mathematical theory: evolutionary game theory. It assumes nothing about the nature of reality. It applies to the evolution of creatures, but also to the evolution of songs, ideas, and scientific theories — it is just as applicable to abstract concepts as to animate beings. This all-inclusive applicability has earned it the nickname "universal Darwinism." Evolutionary game theory negates our assumption that objects, such as DNA, are real, existing even if unperceived. It reveals that these objects are just icons that we create (and destroy) as needed to guide adaptive action.

Isn’t ITP old news? A rock looks solid and dense. But since 1911, when the physicist Ernest Rutherford discovered the atomic nucleus, physicists have told us that a rock is mostly empty space, with tiny electrons darting far from atomic nuclei. Physicists have long known that our perceptions sometimes fail to imitate reality. Indeed, but there is something new in ITP that is profoundly different from the insight of Rutherford. Suppose I admit that a red icon on the desktop is just an icon, not the reality. Then I pull out a magnifying glass, stare closely at the icon, see pixels, and proclaim that those pixels are the reality. I am, of course, mistaken. The pixels are in the screen, still part of the desktop interface. Similarly, tiny nuclei and electrons are in spacetime, still part of our spacetime interface. ITP says that spacetime is not objective reality and does not resemble reality, whatever reality might be. Spacetime is simply a data format that natural selection wired into our species to hide reality and reap fitness payoffs.

The claim that spacetime is not reality is heady stuff. Surely physicists disagree on the subject? In fact, many physicists agree, including Nima Arkani-Hamed, one of our leading theoretical-physicists; he argues that “spacetime is doomed” and that physics must plumb a deeper reality from which spacetime emerges. What this reality might be is unknown, but physicists are pursuing leads that may someday reveal what reality actually is.

¤

I don’t know what reality lies behind our spacetime interface. But I have a theory, motivated by the “hard problem” of consciousness. The problem is this: we find many correlations between brain activity and conscious experiences, such as the taste of salt or the feel of silk. Your experience of color, for instance, is correlated with activity in a region of your brain called area V4; if we stimulate this area electrically then you will experience illusory colors, and if a stroke damages this area then you will lose some of your

normal color experience. We don't know why these correlations exist. The hard problem is to devise a scientific theory that explains these correlations.

This problem is old. The German polymath and philosopher Gottfried Leibniz understood the problem in the early 1700s. The English biologist Thomas Huxley described the problem vividly in 1869: "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the djinn when Aladdin rubbed his lamp in the story."

To this day, science has not dispelled the mystery. Does neural activity cause conscious experiences? Some think so but have no idea how. No neural cause has been proposed for even one conscious experience. Precisely what neural activity causes, say, the taste of vanilla, and precisely how and why does it do so? No one knows.

Are conscious experiences identical to, rather than caused by, neural activity? Some think so but again cannot give even one example. Precisely what neural activity is identical to the taste of vanilla? No one knows.

Why has the hard problem of consciousness remained intractable for centuries despite determined efforts by brilliant scientists? I think the culprit is our assumption that our perceptions reveal a reality that exists even if unperceived. We see neurons when we peer through microscopes. We assume this means that neurons exist even if we don't peer. We further assume that neurons have causal powers, including the power to create consciousness.

But neurons, like all physical objects, are just icons in our interface and have no causal powers. It's a useful fiction, when playing *Grand Theft Auto*, to assume that the steering wheel causes the car to turn. The fiction is compelling because we can *intervene*; we can spin the wheel left and right

and watch the car turn as we predict. But it is a fiction: there is no feedback from wheel to computer. This fiction is harmless for one who just wants to play the game. But for a programmer who wants to understand how the game really works, the fiction is not harmless. Clinging to the fiction precludes a true understanding of the game.

The same is true of neurons. For everyday neuroscience, it is a convenient and useful fiction to assume that neurons exist and have causal powers, even when not perceived. It is harmless to speak of, say, neural activity in area V1 of visual cortex projecting to area V2 and causing changes there in neural activity. But for a researcher who wants to understand consciousness and how it's really related to neural activity, the fiction is no longer harmless. Clinging to this fiction has made the hard problem intractable.

So, instead of proposing that particles in spacetime are fundamental, and somehow create consciousness when they form neurons and brains, I propose the reverse: consciousness is fundamental, and it creates spacetime and objects. I have published with collaborators a mathematical theory of consciousness, which posits that reality is a vast social network of interacting "conscious agents," in which each agent has a range of possible experiences, and each agent can act to influence the experiences of other agents. We call this proposal *conscious realism*. No object within spacetime is itself a conscious agent; spacetime is simply a format for conscious experiences — an interface — employed by agents like us, and physical objects are just icons in that interface. However, when we interact with other conscious agents who might not use a spacetime format for their own experiences, we perceive our interaction in the format of objects in spacetime.

The simplest agents enjoy just two experiences and two actions. When agents interact, they create new agents, more complex and nuanced — like notes forming melodies and symphonies. In the limit, as ever more agents

interact, they compose new agents of infinite capacity, each savoring a limitless variety of experiences and interactions.

With the introduction of these infinite conscious agents, this theory borders on the spiritual. Think of it: all of the various conscious agents who have at some point affected other conscious agents and have contributed to the formation of the current symphony. As a scientist, I can say that we tread on the turf of spirituality not as trespassers but as allies, fellow students of the human condition, offering new tools. We can run simulations, prove theorems, make precise and testable predictions, and, most importantly, *be wrong*. That is the point of precision: to discover, precisely, what's wrong and think about, potentially, how it may be fixed. How else shall we muster real progress on the big questions: What are we? Why are we? What is good? Why do we experience pain?

For any finite agent, the entire network of agents, in its infinity, is incomprehensible. The agent requires an interface to tame and navigate this complexity. Think of Twitter. There are millions of users, billions of tweets, and many trends. How can one possibly grasp it? Easily, if one has a graphical interface that hides complexity and guides exploration of the Twitterverse. Such tools are standard fare in the new era of big data.

In like manner, spacetime and physical objects are simply our tool, our virtual reality, by which we navigate the infinite network of conscious agents. It's not the best tool; evolution finds solutions that are satisficing (just good enough), not optimal. Nor is it the only tool; other agents deploy an infinite variety of different interfaces. And it's just a tool, not the reality, despite our penchant to reify our interface.

¤

Our spacetime interface opens portals into the realm of conscious agents. If you gaze at your face in a mirror, you see skin, hair, and eyes. But you know,

firsthand, that hidden behind your face there pulsates the vibrant world of your conscious experiences — your hopes, dreams, moods, colors, textures, shapes, flavors, aromas, rhythms, timbres, itches, and thoughts. Your visible face is a portal, albeit dim, to the invisible conscious you. Most of you is concealed by the portal, and the rest loses much in translation: a frown does not resemble sadness. When others see your face, they open a genuine, but limited, portal into your conscious world. Which is not to say that your face is conscious. It's not. *You* are conscious. Your face is an icon in the interface of the viewer.

When I see a dog, my portal into consciousness is dimmer. I guess there is enjoyment of a bone and excitement by a squirrel. When I see an ant, my portal is dimmer still; I have little insight into the experiences behind my icon of an ant. With a rock, my portal is opaque; it offers no obvious insight into experiences behind the icon. My interface has, of necessity, finite limits; when it delivers a rock, it cries uncle — similarly, when it delivers atoms and molecules. But we mistake the limits of our interface for an insight into reality: we take atoms and molecules to be fundamental, and because of that we are tempted to build an ontology based on the physical world.

This mistake is harmless — even useful — for many projects in science precisely because they are confined to exploration within our spacetime interface. For much of physics and neuroscience, it's a useful fiction to assume that physical objects exist and have causal powers even when unperceived. It's a convenient shorthand to speak of a cue ball causing an eight ball to careen into a corner pocket or of neural activity in one area of the brain causing activity in another area. This fiction has fostered impressive scientific advances.

But the mistake is fatal for ventures in science not confined to our interface. We saw this in the hard problem of consciousness. We see it also in a classic, but now pressing, problem: can AI become conscious? The standard

answer is that unconscious circuits and software, if endowed with proper complexity and
dynamics, can somehow create consciousness. “Somehow” is a promissory note that, so far, none can pay.

Agents and interfaces offer a new take on this old chestnut. Our spacetime interface has many portals into the realm of conscious agents. We know of one technology that can open new portals: sex, which creates offspring whose bodies are portals. The question for AI becomes this: can we rejig our interface, using silicon and software and such, to open new portals into the preexisting realm of conscious agents? It’s not a question of creating consciousness, but of forging new access to preexisting conscious agents. For what it’s worth, I think we can. If we succeed in opening new portals, it’s hard to predict what will greet us, given the infinite variety of conscious agents. Perhaps comity, perhaps chaos.

We also get a new take on a famous question posed by the physicist Enrico Fermi: where is everybody? By some estimates, intelligent life should be abundant and evident throughout our galaxy. It’s not. The search for extraterrestrial intelligence has so far come up empty.

Why? According to ITP, we aren’t searching the right place. We search outer space and assume that ETs must lurk there. But space is not reality; it’s our virtual reality, the idiosyncratic interface of our species. We search with our (metaphorical) headsets on. We see what our headsets allow, which isn’t much. That’s the point of our headset: to hide reality and give us icons in spacetime to navigate reality despite our ignorance. Where is everybody? According to conscious realism, they’re out there. But it’s no surprise if they, like dark matter and dark energy, escape our interface. Almost everything escapes it.

¤

Some 2,500 years ago, we let go of flat Earth. Then 400 years ago, we let go of a geocentric universe. It was hard. We burned dissenters at the stake. It really looked to us as though the Earth was the center of the universe. We took our senses too literally.

These were warm-ups. Now we must let go of spacetime itself. We suppose that the long sweep of spacetime, with its countless stars and planets, is the preexisting stage for an accidental drama in which we are bit players. We think it's faintly mad to suppose otherwise. But we're mistaken. We are the authors of space and time; their myriad contents are our impressive stagecraft.

We, like the ancients, believe that perception is imitation — that we see reality as it is, not virtual realities as we need. We have it backwards: natural selection has endowed us with a headset. Heretofore we haven't seen it, because we see through it.

Can we free our minds and recognize our headset for what it is? No doubt yes, thanks to advances in technology. With satellite photographs, one need be no Copernicus to recognize earth's place in space. Soon, countless virtual worlds — for business and entertainment — will be as common, compelling, and immersive as daily life. The idea that *this* reality, with headset off, is somehow uniquely veridical, will prompt the same curious amusement as flat Earth.

The insight of Copernicus opened, in due course, a window to modern space travel. Where will our imagination take us, once we free our minds?

¤

Donald D. Hoffman is a professor of cognitive sciences at the University of California, Irvine and the author of The Case Against Reality: Why Evolution Hid the Truth from Our Eyes *(2019).*