COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2022) e13195

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13195

Flexible Goals Require that Inflexible Perceptual Systems Produce Veridical Representations: Implications for Realism as Revealed by Evolutionary Simulations

Marlene D. Berke, Robert Walter-Terrill, Julian Jara-Ettinger, Brian J. Scholl

Department of Psychology, Yale University

Received 18 November 2021; received in revised form 3 August 2022; accepted 5 August 2022

Abstract

How veridical is perception? Rather than representing objects as they actually exist in the world, might perception instead represent objects only in terms of the utility they offer to an observer? Previous work employed evolutionary modeling to show that under certain assumptions, natural selection favors such "strict-interface" perceptual systems. This view has fueled considerable debate, but we think that discussions so far have failed to consider the implications of two critical aspects of perception. First, while existing models have explored single utility functions, perception will often serve multiple largely independent goals. (Sometimes when looking at a stick you want to know how appropriate it would be as kindling for a campfire, and other times you want to know how appropriate it would be as a weapon for self-defense.) Second, perception often operates in an inflexible, automatic manner—proving "impenetrable" to shifting higher-level goals. (When your goal shifts from "burning" to "fighting," your visual experience does not dramatically transform.) These two points have important implications for the veridicality of perception. In particular, as the need for flexible goals increases, inflexible perceptual systems must become more veridical. We support this position by providing evidence from evolutionary simulations that as the number of independent utility functions increases, the distinction between "interface" and "veridical" perceptual systems dissolves. Although natural selection evaluates perceptual systems only on their fitness, the most fit perceptual systems may nevertheless represent the world as it is.

Keywords: Computational modeling; Evolution; Interface theory of perception; Realism; Visual perception

Correspondence should be sent to Marlene Berke or Brian Scholl, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520–8205, USA. E-mail: marlene.berke@yale.edu, brian.scholl@yale.edu

1. Introduction

Do we see the world as it truly is? On one hand, we are all familiar with cases in which perception is inaccurate—sometimes to a striking degree, as in visual illusions. (You may know with certainty that two lines in front of you are equally long, while still irresistibly seeing one as longer than the other, as in the Müller–Lyer illusion.) But on the other hand, most of us still intuitively assume that these are the exceptions that prove the (opposite) rule: our percepts reflect the external world faithfully, and are useful only to the degree to which they are accurate. (If you see a bunny in front of you, but it is really a tiger, you may pay the price for that inaccuracy.) Recently, however, a provocative view has emerged that questions such assumptions, arguing that our perceptual representations of the world are in general very different from the actual state of the world.

1.1. A new evolutionary perspective: The "interface" theory of perception

The "interface theory of perception" (henceforth ITP) proposes that our percepts may almost always differ radically from the ground-truth to the extent that even key dimensions of perception (such as space and time) may have little to no basis in external reality (Hoffman 2009, 2014, 2016, 2018, 2019; Hoffman & Prakash, 2014; Hoffman, Singh, & Prakash, 2015; Mark, Marion, & Hoffman, 2010). According to ITP, illusions are the rule, and veridicality[1] a rare exception.

1.1.1. Conceptual introduction

ITP is inspired by straightforward evolutionary considerations. Perception can only make a difference—and can only have evolved in the first place—if it impacts our fitness. And so, we should therefore not think of perceptual representations in terms of their correspondence to objective external reality, but rather to the underlying utility functions that relate behavior to fitness. If reality differs in some way that is orthogonal to utility, then those differences would have no impact on our fitness, and there is no mechanism by which we could have evolved an accurate representation of those aspects of reality. And conversely, if distinctions that do not reflect an underlying reality (e.g., between space and time) nevertheless have fitness benefits, then they will become incorporated into our percepts. In short, whenever an element of the world differs in terms of its objective properties and its subjective utility, the most fit perceptual system will always be one that represents the element in terms of its subjective utility.

To help make this perspective clear, consider a concrete example. Imagine a frog that lays eggs in pools of brackish water, and that the eggs' viability depends in part on the water's salinity: if the frog lays its eggs in a pool that is either too salty or insufficiently salty, then the eggs will not hatch. And assume that the function relating salinity to viability is both unimodal and symmetric, centered on some ideal value (i.e., a bell curve). That function might be represented by the generic utility curve depicted in Fig. 1a (adapted from Hoffman et al., 2015), where the horizontal axis represents salinity, and the vertical axis represents the eggs' viability. Here, the four colored regions represent different percepts that the frog could

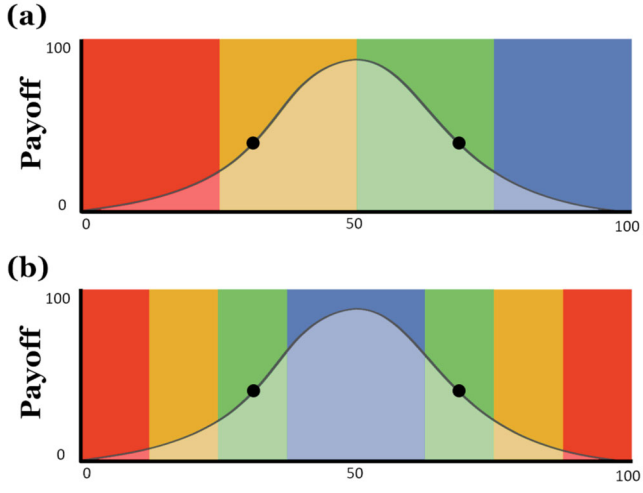

Fig. 1. (a) A hypothetical veridical perceptual system; and (b) a strict-interface perceptual system. In each case, the horizontal axis represents an objective dimension of the world (e.g., the salinity of pools of water), and the vertical axis represents the evolutionary payoff of each value of that dimension (e.g., the viability of a frog's eggs). The curve then depicts a hypothetical utility function relating the underlying dimension to its payoff, and the colors indicate four possible binned percepts (e.g., of different levels of salinity). In the veridical perceptual system, the two values highlighted by the dots would correspond to different percepts (of different absolute salinity levels). In the strict-interface perceptual system, these two values would correspond to the same percept, since they have equivalent utilities. (Adapted from Hoffman et al., 2015).

have—distinguishing different binned degrees of salinity. But according to the logic of ITP, there is no mechanism by which this perceptual function could evolve. Consider, for example, the two highlighted points on this curve: these points correspond to two objectively different salinity levels, but those salinity levels have the very same impact on the eggs' viability, and so there would be no utility to distinguishing between them. Instead, ITP proposes that what would evolve would necessarily be a perceptual function such as the one depicted in Fig. 1b—wherein these same two points correspond to identical percepts (here depicted by the fact that they both lie within the green regions). In this toy example, the frog's resulting perceptual system would essentially be representing the "goodness" or "badness" of the salinity levels (and not their absolute values), since only the former would have any downstream fitness consequences.

This perspective was dubbed an "interface theory" because of its metaphorical connection to other familiar types of technological interfaces. In ITP's analogy, a computer desktop represents the underlying information in the computer in terms of folders, icons, and perhaps a trash bin—but these are "useful fictions" to a large degree: two files that appear inside the same folder, for example, may actually be stored in entirely different (or even fully intermingled) locations in the underlying hard drive. (And this example makes clear why this particular fiction is useful—since a desktop that explicitly depicted the actual locations in memory would be distracting, useless, or worse.) Similarly, ITP suggests that much (or all)

of human perception is a useful fiction—nothing but a user "interface" to the outside world, which may radically distort (or even hide) external reality.

1.1.2. Computational support

Theoretical debates about these questions of "perceptual realism" are legion, and have been heavily featured both in the history of philosophy and in contemporary philosophical work—for example, contrast the realist views of Campbell and Cassam (2014), Locke (1690), or Searle (2015) with the antirealist arguments of Berkeley (1709, 1725), Jackson (1977), or Robertson (1994). And more recently, these debates have featured evolutionary arguments on both sides (e.g., Korman, 2019; McKay & Dennett, 2009; Popper, 1987; Shepard, 1990, 1992, 1994, 2001; Thompson, 1995). But the ITP has had a notable impact on recent discussions in part because it has taken an entirely new (and fresh) approach to the thoughts and arguments reviewed above by supporting them with data from evolutionary simulations using "genetic algorithms."[2]

Genetic algorithms are a method in computer science and artificial intelligence that is useful for understanding how computational systems are shaped by evolutionary pressures. In genetic algorithms, a set of artificial agents (often generated randomly) are placed in a simulated environment and must complete a series of decisions that yield different payoffs. In the context of ITP, the environment consists of a collection of resources, each associated with a different payoff, and each agent is initialized with a random perceptual system (i.e., an arbitrary mapping from resources to perceptual representations). Each agent is then allowed to make a sequence of choices about which resources to "forage" and its final payoff is given by the sum of payoffs obtained.

After simulating how the initial set of artificial agents behaves in the simulated environment (perhaps across many rounds of foraging), a new generation of artificial agents is then created through a process of selection, recombination, and mutation. In this step, agents are selected for reproduction as a function of their fitness, such that agents who performed better in the simulated environment have a higher chance of passing on their cognitive system (and, conversely, low-fitness agents have a low chance of passing down their cognitive system). New agents are then generated by combining the cognitive systems of pairs of selected agents, introducing a small probability of mutation which allows for the introduction of novel cognitive systems. In the context of ITP, this means that the perceptual strategies of agents that selected better resources will be represented in the next generation of agents, while the perceptual strategies of agents that made poor choices will die off. By repeating this process over many (perhaps hundreds, or even thousands, of) generations, those perceptual strategies with the highest fitness scores will end up dominating the population. And correspondingly, one can then look at the distribution of strategies that remain after these simulations are complete in order to identify those that are the most fit. A crucial benefit of this type of simulated evolution is that this process can produce entirely new (and better-adapted!) strategies that were not even present in the initial population.

These are the sorts of evolutionary simulations that were conducted with veridical versus "strict-interface" perceptual strategies (initially by Hoffman, 2009; Mark et al., 2010; later published and reviewed in Hoffman et al., 2015). (While an "interface perceptual strategy"

could accidentally be a homomorphism of real-world structures, a "strict-interface perceptual strategy" is one that lacks any coincidental homomorphism to the real world.) And in this context, the results were clear and compelling: strict-interface perceptual strategies (such as that depicted in Fig. 1b) consistently dominated in the final population, while the veridical perceptual strategies (such as that depicted in Fig. 1a) effectively died out. This is the empirical/computational support for ITP's suggestion that in a contest between strategies tuned to objective reality and subjective utility, the latter always wins.

1.2. The current project: Flexible goals and inflexible perceptual systems

ITP has gained widespread attention, having been recognized as an unusually distinctive and influential contribution. For example, one of its recent incarnations (Hoffman et al., 2015; see also Hoffman, 2009, 2014; Hoffman & Singh, 2012) so impressed the Editor of *Psychonomic Bulletin and Review* that it was not only published but was also featured in its own special section, with a breathless introduction (questioning whether it might be "the future of the science of the mind"; Hickok, 2015, p. 1479), along with associated commentaries from 10 other groups and an extensive reply by the authors. Given how provocative ITP is, it is perhaps no surprise that it has also been criticized on several grounds—for example, for its (mis)characterization of veridicality (Cohen, 2015; Edelman, 2015), for its failure to consider key roles of adaptation on ontogenetic timescales (Anderson, 2015), for ignoring other experimental evidence in favor of veridicality (Pizlo, 2015), and for failing to explain how interface strategies could be functional in novel situations (Jansen, 2018).

Our goal in the present paper is not to review this entire (large, multidimensional) debate (see also Angelucci, Fano, Ferretti, Macrelli, & Tarozzi, 2021; Feldman, 2015; Fields, 2015; Koenderink, 2015; Martinez, 2019; Mausfeld, 2015; McLaughlin & Green, 2015; O'Connor, 2014; Schlesinger, 2015; Wilson, 2021), however, and we will largely refrain from weighing in on these many previous critiques and subsequent defenses. Instead, we aim to contribute something new to this discussion, by presenting some key considerations concerning our "cognitive architecture" that pose a stark challenge to ITP (and support the intuitive notion of veridical perception)—but that to our knowledge have not previously been considered in past discussions (by either ITP's proponents or detractors). In particular, we seek to explore the implications of two ideas, both related to (in)flexibility: (1) perception must be flexible enough to serve multiple largely independent goals, but (2) the cognitive architecture of perception often renders it "informationally encapsulated" and "cognitively impenetrable"—and thus unable to flexibly change in response to shifting goals. In this section, we will expand on the nature and theoretical implications of these two ideas, and then in Sections 2 and 3, we will explore their computational implications when they are implemented in the context of evolutionary games and genetic algorithms.

1.2.1. Flexible perceptual goals

In the concrete example discussed above, the frogs need to perceive salinity in order to assess the viability of the water for laying eggs. But of course, that may not encompass all of the frogs' needs and goals. Suppose, for example, that the same frogs also need to eat fly

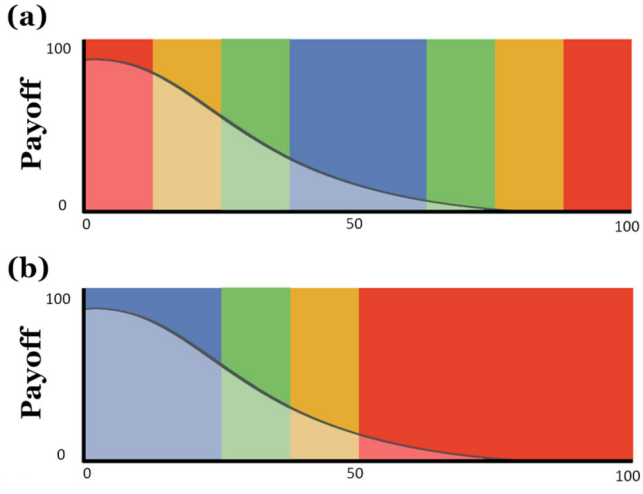

Fig. 2. A hypothetical utility function (as indicated by the curves) that prioritizes the ability to detect low values. (a) When the same interface perceptual system from Fig. 1b (as indicated by the colored bars) is matched with this function, the fit is poor. (b) But this same utility function is well-matched to a very different interface perceptual system.

larvae, and the most nutritious fly larvae tend to be found in freshwater pools. Clearly hungry frogs which are able to determine which pools contain freshwater will then accrue fitness benefits, and a veridical perceptual system (such as that in Fig. 1a) would allow for this. But this would prove deeply problematic for a frog whose strict-interface perceptual system is tuned to egg-laying and which thus cannot distinguish between pools of high versus low salinity (as in Fig. 2a, where the same perceptual system from Fig. 1b is now matched with an ill-fitting utility function). And similarly, a frog with a perceptual system that *was* only able to distinguish between especially low salinity and everything else (such as that depicted in Fig. 2b) might be well adapted for finding food, but that same frog would then be at a stark disadvantage when needing to lay its eggs. Put differently, the two perceptual systems depicted in Figs. 1b and 2b are each individually well-adapted to a particular goal, but they are irreconcilable with each other—and so to meet both goals at the same time requires a more multipurpose system, moving the frog closer to something like Fig. 1a.

These ideas can be generalized: perception will often support behavior in the context of multiple largely independent goals. Sometimes when looking at an apple, for example, you may want to know how appropriate it is for eating, and in that context, it may be helpful to analyze its properties (say, its specific spatial luminance profile) in one way, as a cue to ripeness. But other times—for example, when your prospective snack is rudely interrupted by the appearance of a predator—you may suddenly want to know how appropriate the apple is for throwing, and in that context, those same properties may suddenly need to be analyzed in a very different way, as a cue to solidity or density. (This sort of flexibility has often been noted in the context of "affordances," insofar as most objects afford many disparate actions; Gibson, 1977.) In particular, the function that associates utility with different regions of the relevant

luminance space might be very different in the context of these different goals. A person with a strict-interface system that is well-adapted to eating might be less adept at choosing projectiles, and vice versa—while a person with a veridical perceptual system might excel at both. And critically, these two functions could very well be in effective conflict: the portion of a stimulus space that conveys the most fitness under one goal might convey the least fitness under a different goal.

This perspective has been obscured in most past discussions of ITP, which have typically only engaged with a single task or utility function at any given time. (In one of the earliest ITP papers, Hoffman & Singh [2012] did acknowledge that organisms may not always have only one utility function, but speculated that "there is no principled reason why" more fitness functions would necessarily produce veridical perceptual strategies, and noted that "this must remain an open question until detailed mathematical models of this process are developed and studied" [p. 1086]. However, such models were not addressed in subsequent work. And in the large and growing body of commentary on ITP, we know of only two previous papers that have considered this factor, albeit from rather different perspectives; cf. Angelucci et al., 2021; Martinez, 2019.) But surely having multiple independent potential goals is the norm for human perceptual systems. Indeed, the toy examples discussed above with frogs and apples surely vastly understate the extremity of this variability, as we shift so frequently among so many different goals—related to food, mating, child-rearing, predation, clothing, shelter, politics, war, entertainment, and on and on—in our everyday experience. (To make this point more concrete, consider all the many and diverse goals you might have when looking at a tree branch—e.g., assessing its utility as a potential weapon, as kindling for a fire, as a walking stick, as a fishing pole, as a backscratcher, as a tool for knocking an apple out of a tree, as a patch for your canoe, as a support for your wigwam, as a drumstick, as a splint for a broken bone, as a flagpole, as a fencepost, as an oar for rowing, as a mast for a sail, as something to sit on, as something to artistically carve into, or as a stick for roasting marshmallows.) And this does not even yet include the development of entirely new goals and utility functions; for example, part of the majesty of human perception and cognition is our ability to readily adapt to entirely novel goals that may have never faced past generations. (Perhaps you want to use your tree branch as a replacement leg for your standing desk?)

All of this seems important, insofar as the conclusions of ITP may seem far more intuitive and compelling in the context of individual, isolated goals and utility functions, compared to the multiplicity of goals alluded to above. And the key point here is not that any of this obviates the critical role of fitness in shaping our perceptual systems: this underlying logic of ITP seems correct and undeniable. Of course, that logic can also still apply just as readily even in the face of a multiplicity of goals: one can just average all of these utility functions together, and then point out again that what we perceive may reflect only this global/averaged utility function, rather than any objective external reality. But what matters for an organism's fitness is not some average utility of a resource (e.g., an apple) across all possible contexts, but rather the utility of the resource at each moment the organism has to act (when feeding, or fending off predators). The conclusions of ITP may thus be undercut, since although the resulting perceptual system may still technically be an interface, it may also be veridical. Indeed, we suspect that as we allow more and more independent (and potentially conflicting)

goals (and corresponding utility functions) into the mix, the distinction between an interface and veridical systems will simply collapse—since, in the end, the perceptual system that is most fit (and most flexible) in the context of many shifting goals will be a veridical one.

1.2.2. Inflexible (and impenetrable) perceptual systems

The lesson of the previous section was that perception needs to be *flexible*, insofar as it must serve many different shifting goals (and corresponding utility functions), and that when these goals and utility functions are all considered in tandem, the resulting interface systems converge on veridicality. But one way that ITP could avoid such implications is simply to deny that there is any such "global" consideration in the first place: instead of fitness operating through a single global "average" utility function, different local utility functions could just be considered *sequentially*, with only one operating at any given moment to determine the character of what we see. When you want to eat an apple, it may look one way—but when you want to throw it, it may appear entirely different. This would be a way to salvage the provocative conclusions of ITP: rather than looking at one (mostly veridical) way all of the time, objects would appear in many different (and differently nonveridical) ways from goal to goal, and from moment to moment.

Our second point is simply that actual human perception does not seem to work like this since the underlying "cognitive architecture" of perception does not allow for goals to influence either the details of visual processing or the resulting percepts in this way. Instead, perception is "cognitively impenetrable" due to a form of informational encapsulation: a particular process (or "module") in the visual system, for example, may have access to certain inputs (often in the form of the shifting patterns of light on the retinae, and the subsequent representations of the environmental factors that are deemed via unconscious inferences to have caused those patterns of light), but the process may *not* have access to information about conscious or deliberative goal states from other parts of the mind—and accordingly, there is no way for those goal states to change the nature of the processing or its resulting percepts, aside from directing attention to some properties more than others. This is a familiar phenomenon from almost all visual illusions (for discussion, see van Buren & Scholl, 2018): once you learn that the illusion is in fact an illusion (and perhaps even understand exactly how it arises), that newfound knowledge and certainty does nothing to diminish the perceptual oomph of the illusion itself. (You continue to *see* the two lines in the Müller–Lyer illusion as having different lengths, even while you *know* that they are the same length—say, because you just measured them with a ruler.)

This perspective has been supported by a wealth of both theoretical arguments and empirical studies. Theoretically, it has been argued that although certain distortions of perception by high-level states could be helpful in particular circumstances, those same distortions might be highly maladaptive in other circumstances—and you cannot always know in advance just which circumstances you may suddenly find yourself in. If wearing a heavy backpack makes hills look steeper, for example (e.g., Bhalla & Proffitt, 1999), then that could be helpful in leading to wiser decisions about whether to climb them in some circumstances (say, in terms of conserving energy)—but it could also lead to devastating decisions in other circumstances (e.g., making the hill seem more safe than it actually is during a flood). This has recently

been phrased in terms of the same motivations as a "free press" enjoys (Gilchrist, 2020): top-down distortions of news by a government could certainly be useful for certain political goals, but "if the information is distorted, the well is poisoned, with serious damage to other functions, educational and scientific, among many others, that depend on reliable information" (p. 1002). So too with perception: just as societal functioning may result in pressure for accurate news reporting, evolutionary pressures may drive perception toward veridicality.

Empirically, there have been hundreds of reports of putative influences of top-down states (such as conscious, voluntary goals) on perception, but later work has often demonstrated that these apparent effects are better explained away in other ways. The putative influence of heavy backpacks on perceived hill steepness, for example, has been shown to depend on task demands—such that if you provide an alternate explanation for why participants are wearing a backpack (of the same heaviness), the effects vanish (Durgin et al., 2009; see also Firestone & Scholl, 2014). And more generally, it has been argued that these hundreds of purported effects can be collectively deflated by only a small handful of common empirical "pitfalls"—such as failing to recognize subtle low-level visual differences that are correlated with the high-level factors (e.g., Firestone & Scholl, 2015a), mistaking effects on memory for effects on visual experience (e.g., Firestone & Scholl, 2015b), or mistaking higher-level judgment for perception (e.g., Woods, Philbeck, & Danoff, 2009). In each case, critiques of the initial studies not only argued that such pitfalls *could* explain the relevant top-down effects in principle, but they also showed empirically that those pitfalls *did* actually explain those effects, in practice. (And more generally, researchers have tended to look only for confirmatory evidence in support of such top-down effects, rather than checking to make sure that such effects also *do not* appear when the "cognition affecting perception" theory says that they should not; e.g., Firestone & Scholl, 2014.)

Our goal in the present paper is not to mount a defense of the cognitive impenetrability of perception since this view remains deeply controversial, and since the controversy has been extensively documented elsewhere (for an empirically oriented review with many vigorously objecting commentaries, see Firestone & Scholl, 2016). Rather, our argument here is a conditional one: *if* perception is cognitively impenetrable (as many have argued), *then* it is not possible for proponents of ITP to appeal to the possibility that each different goal activates a different perceptual interface, which in turn drives a different experience (each of which may individually diverge dramatically from reality).[3] And to our knowledge, no previous discussions of ITP have ever acknowledged this connection to the (im)penetrability of perception.

1.2.3. Putting the pieces together: Inflexible perceptual systems supporting flexible goals

Cognitive impenetrability on its own poses no direct challenge to ITP—since we could readily perceive the world via a single (unpenetrated) interface of the sort described in Section 1.1.1 (along the lines of Fig. 1b). And the need to flexibly serve multiple goals on its own poses no direct challenge to ITP—since our experience could just shift among different radically false interfaces whenever our goals shift. But we suggest here that when these two points are put together, ITP cannot survive: when an inflexible perceptual system must flexibly serve many goals, the distinction between veridical and interface systems dissolves—as the best interface becomes one that is veridical.

2. Computational modeling

The arguments in the previous section contribute to a long-running theoretical conversation about the possibility of perceptual realism, but ITP has become prominent less for its own arguments and more for the computational models which showed just how interface perceptual systems "win" in evolutionary contests, driving veridical systems to extinction. In the present project, we embrace this approach and explore the computational consequences of the two primary points from Sections 1.2.1 and 1.2.2, when implemented in the framework of genetic algorithms. In particular, we adopt the evolutionary modeling framework that ITP introduced, and we show that expanding the number of tasks (per Section 1.2.1, operationalized as distinct payoff functions) has a dramatic influence on the perceptual strategies that "win" such evolutionary contests—at least when perceptual strategies cannot change from one payoff function to the next (per Section 1.2.2). When there is only one (or a few) task, we replicate previous ITP findings that the most fit strategy is usually a nonveridical interface tuned to that particular task. But as the number of independent tasks increases, the most fit strategies become increasingly veridical.

2.1. Elements adopted from previous work

We first highlight the high-level similarities between ITP simulations and ours, before discussing the implementational details specific to our work. For the current model, we adopted the ITP framework of evolutionary games as described in Section 1.1.2, including individual agents with differing perceptual strategies making choices between resources that offer different fitness payoffs. After many rounds per generation (during which these perceptual strategies stayed fixed), these agents then reproduce based on their accrued fitness, with strategies reproducing—with random mutation—at generational breakpoints, and producing gradual evolution of perceptual strategies across hundreds of generations.

Beyond these core components of genetic algorithms, our model also adopts many more particular elements from the previous modeling work that has been used to support ITP (Mark, 2013), including: (1) the space of resources; (2) the definition of payoff functions mapping resources to fitness payoffs; and (3) the definition of veridical, nonveridical, and interface perceptual strategies.

As in previous models, we use 11 possible resources {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11}, and these resources have a ground-truth ordered structure ($1 < 2 < 3 < \ldots < 11$). A payoff function maps each resource to the fitness payoff for acquiring that resource. An example of a payoff function could be {0, 2, 10, 15, 19, 20, 19, 15, 10, 2, 0}, where resource 1 offers a payoff of 0, resource 2 offers a payoff of 2, resource 3 offers a payoff of 10, and so on (here with the maximum payoff of 20 being offered by resource 6). When an organism chooses a resource, it gets the fitness payoff corresponding to the resource it chose under the payoff function.

A perceptual strategy is defined as a mapping from the resources to perceived colors. Using two colors—"r" for red and "g" for green—a perceptual strategy could be {r, r, r, r, r, r, g, g, g, g, g}, or {6r, 5g} for short. This perceptual strategy maps resource quantities of 1–6 to

red and 7–11 to green. Perceptual strategies are classified as veridical if there is an order-preserving homomorphism between the resource in the real world and the perceived color, such that all resources producing the percept of red were earlier/smaller (or later/bigger) in real-world structure than all resources producing the percept of green. Or equivalently: a veridical strategy is one without any disjoint between the perceived colors. Veridical strategies are thus those like {2r, 9g} or {5g, 6r}, but not those like {3r, 6g, 2r} or {1r, 1g, 1r, 1g, 1r, 1g, 1r, 1g, 1r, 1g, 1r}. Nonveridical perceptual strategies are simply those that are not veridical under this definition. And interface strategies are those that reflect the ordering of the resources according to payoff. So, given a payoff function like {0, 2, 10, 15, 19, 20, 19, 15, 10, 2, 0}, interface strategies are strategies like {3r, 5g, 3r} or {2g, 7r, 2g}.[4]

2.2. Novel properties of the current models

From the perspective of the current project, the key property of the previous ITP models was that agents only ever had to perform a single task—or only had to make choices under a single payoff function. In our simulations, in contrast, agents complete several different tasks during each round—operationalized as choosing under different payoff functions—and fitness results from payoffs accumulated across many tasks with different payoff functions. We run these evolutionary simulations across several conditions. In the simplest condition (most similar to the original ITP simulations), there is only one payoff function: each individual performs 100 rounds of making decisions under the same payoff function, performing the same task 100 times. In other conditions, however, we use multiple payoff functions (ranging from 2 to 2000)—where each round employs a payoff function that is sampled with replacement from the set of possibilities. None of these payoff functions are monotonically increasing (or decreasing) functions, as under monotonic payoff functions, interface strategies will trivially be veridical (Hoffman et al., 2015). Additionally, the agent always knows the current relevant payoff function. Our models also eliminated recombination, and simply had each parent produce a copy of itself with mutation. We implemented this form of asexual reproduction simply because it is more likely to preserve the properties of the parental perceptual systems from one generation to the next.

2.2.1. Implementation details and parameters

Specifically, each simulation begins with an initial population of 1000 agents, each of which has a perceptual system that is created by sampling from a uniform distribution over all possible permutations. All agents then perform the same 100 rounds. During a round, each agent is presented with a set of 2–11 resource options (with the set size sampled from a discrete uniform distribution from 2 to 11, and the resources are sampled without replacement from the set {1, …, 11} of all resources) and must choose one of them. Critically, as an implementation of cognitive impenetrability, the agent must always use the same mapping from stimulus (resource) to percept (color), regardless of what the current payoff function is. At the same time, actions are allowed to vary depending on the goal context. Using the current payoff function, an agent calculates whether red or green resources have a higher expected payoff, and then simply chooses a resource of that color.

Performance on these tasks determines reproductive success. After each agent has performed these 100 rounds, its total fitness is calculated by summing the payoffs of each resource that it chose. Each agent then probabilistically produces a number of offspring proportional to its fitness (with parents sampled from the current populations of agents according to a multinomial distribution where $n = 1000$ and $p_i = \frac{fitness_i}{\sum_{j=1}^{1000} fitness_j}$). To produce offspring from a parent, the parent's perceptual strategy is copied, but each of the 11 colors in a strategy ("r" or "g") independently has a 0.001 probability of switching to the other color.[5] The offspring from this process then participate in the next generation, whereupon this full process restarts. In our primary set of simulations—aimed at exploring the effect of the number of tasks on the resulting perceptual strategies—we ran each simulation for 1500 generations, and ran 1000 simulations for each of nine different numbers of payoff functions (1, 5, 25, 100, 200, 300, 400, 500, and 2000), for 9000 simulations in total. In our secondary set of simulations—aimed at exploring the effect of the similarity between two tasks on the resulting perceptual strategies—we ran each simulation for 500 generations, and ran 100 simulations for each of 90 different bins of overlap between a pair of payoff functions ([0.10–0.11], [0.11–0.12], …, [0.99–1.0]), for 9000 simulations in total.

The payoff functions are beta functions that are then discretized, such that the payoff of a resource n is the integral of the beta distribution over the interval from ((n–1)/11, n/11). These beta functions are sampled based on their mode and concentration (following the reparameterization of beta distributions from Kruschke, 2015, p. 129), with the mode sampled uniformly on (0,1) and the concentration sampled according to an exponential distribution (with lambda = 1/15 in the primary set of simulations, and lambda = 1 in the secondary set of simulations).[6] In both sets of simulations, monotonic functions were discarded and replaced.

3. Results

3.1. Veridicality as a function of the number of tasks

The central result from our primary simulations is that as the number of possible tasks increases, the most fit perceptual strategy becomes veridical.[7] This pattern is depicted in Fig. 3. When there is only one task, operationalized here as a single payoff function (as depicted here by the red line at the bottom of the graph), the vast majority (92%) of simulations result in a nonveridical perceptual system—with this result thus largely replicating the results of the original ITP simulations, where veridical perceptual strategies effectively died out. The 8% that do still evolve veridical perceptual systems often look something like {3r, 8g} (that, while not disjoint, do not carve up the resource space evenly).[8] These kinds of strategies tend to evolve when the underlying payoff function is higher at one end than at the other. (Although none of the payoff functions used were monotonic, sometimes the highest valued resources are ones grouped at one end—so that the interface strategy of distinguishing high-valued resources from low-valued ones just happens to preserve the ground-truth. If we do not allow such "accidentally" veridical strategies to count as veridical, then the red line

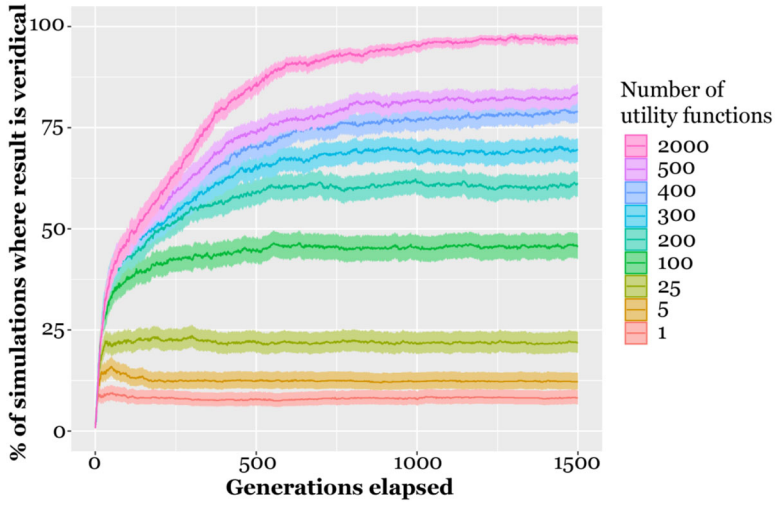

Fig. 3. The central results from our primary evolutionary simulations, exploring the connection between the number of tasks and the evolution of veridicality. The *x*-axis is the number of generations that have been simulated, and the *y*-axis is the proportion of simulations where the dominant (most common) strategy is a veridical one. Each colored line shows the results for a condition with a different number of tasks (payoff functions). The colored error bars represent 95% confidence intervals.

will effectively drop to 0% of simulations resulting in veridical strategies, as found in the initial ITP simulations.)

As the number of tasks and payoff functions increases, so too does the proportion of simulations in which the dominant strategy ends up being veridical—where these perceptual systems tend to be those that evenly divide up the resource space, such as {5r, 6g}.[9] Thus, we can see that (1) considering only five payoff functions (rather than just one) effectively increases the propensity of veridical strategies to dominate in a simulation by 50% (from 8% to 12%, as depicted by the difference between the red and orange lines [i.e., the two lowest lines] at the rightmost edge of Fig. 3); (2) by the time we consider 200 payoff functions, the strategies that evolve and dominate are veridical on the majority of simulations (as depicted by the turquoise line [fifth from the top] in Fig. 3); and (3) by the time we consider 2000 payoff functions, the nonveridical interface theories have been overrun on nearly every simulation (as depicted by the pink line near the top of Fig. 3, where 97% of the simulations result in perceptual strategies that end up being veridical).

Beyond considering the dominant strategy, we can also examine the *distribution* of strategies that evolve. In the condition with 2000 utility functions, in the final generation, there are 1000 agents per each of the 1000 simulations, resulting in 1 million total agents. We found that 82% of those agents have either {5g, 6r}, {6g, 5r}, {6r, 5g} or {5r, 6g} as their perceptual strategy, with each of those four strategies equally prevalent. Other veridical strategies make up another 2%. The most common interface strategies differed from veridical strategies only by one letter at the end: {6g, 4r, 1g}, {1g, 4r, 6g}, {6r, 4g, 1r}, {1r, 4g, 6r}, {5r, 5g, 1r}, {5g, 5r, 1g}, {1r, 5g, 5r}, and {1g, 5r, 5g}, collectively totaling to 11% of the total. The remaining

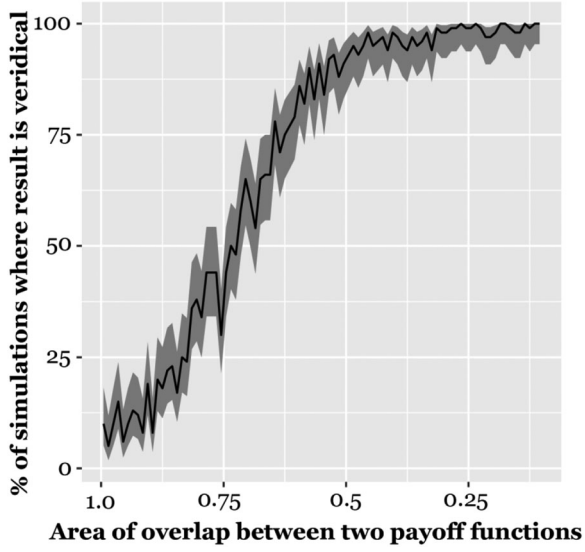

Fig. 4. The central result of our secondary evolutionary simulations, exploring the connection between task similarity and the evolution of veridicality. The horizontal axis represents the overlap between the two utility functions, and the vertical axis represents the proportion of simulations where the dominant (most common) strategy was a veridical one. The error shading represents 95% confidence intervals.

5% of strategies were assorted interface strategies (e.g., {1r, 1g, 3r, 6g}). While veridical perceptual systems are highly prevalent (> 80%) under 2000 utility functions, they do not even make it into the top 10 strategies under one utility function.

3.2. Veridicality as a function of the variety of tasks

The primary analyses in the previous section clearly showed a powerful role for the number of tasks, but those analyses may also seem to suggest that veridicality will not evolve unless or until there are many (possibly hundreds, or thousands of) distinct tasks. But in fact, it is not just the number of tasks that matter, but also how distinct those tasks are—as we explored in a set of secondary simulations, finding that even just two highly distinct tasks can cause selective pressure for a perceptual system to become veridical. One way to operationalize the similarity between two tasks is by using the overlap of the two payoff functions. This relationship between the similarity between two tasks and the resulting evolution of veridicality is depicted in Fig. 4. When the two tasks are nearly identical (as in the leftmost part of the plot)—operationalized here as two payoff functions with very large areas of overlap ($0.99 <$ area of overlap < 1.0)—then the vast majority (90%) of simulations result in a strict-interface perceptual system. This result largely replicates the original ITP simulations, finding that veridical perceptual strategies do not evolve when there is effectively only one task. But when the two tasks differ tremendously from one another (as in the rightmost part of the plot)—operationalized here as two payoff functions with small areas of overlap ($0.10 <$ area of overlap < 0.11)—then veridical perceptual systems dominate (with 100% of simulations

resulting in a veridical perceptual system). And even when the two tasks are only moderately different from one another ($0.50 <$ area of overlap < 0.51), veridical perceptual systems readily evolve (with 91% of simulations resulting in a dominant veridical perceptual system). These simulations demonstrate that even just two highly distinct tasks can cause selective pressure for a perceptual system to become veridical.

4. Discussion

ITP holds that perceptual systems which are tuned to the real world (and are thus veridical) will inevitably be less fit than perceptual systems tuned directly to a payoff function. And when tested across numerous evolutionary simulations—always with just a single payoff function—the result was always the same: interface perceptual systems thrived, while veridical perceptual systems almost always went extinct. This central modeling result helped ITP to transcend past theoretical discussions and has fueled most of the excitement and controversy surrounding this view. And in the present work, while adopting this modeling approach, we readily replicated this result when our simulations were also limited to a single payoff function.

Everything changed, however, when we tweaked the nature of these simulations in two related ways—to allow for both more and less flexibility. First, we embraced the need for perception to accommodate a wide range of goals—implemented in our simulations by introducing different underlying payoff functions for different choices, with these different functions all operating within the same simulated generation. Second, we adopted a form of "cognitive impenetrability" within this framework, such that switching from one goal or task to another could not effectively make the world appear entirely different—implemented in our simulations by requiring each agent to have a *single* perceptual system that helps to determine each choice, regardless of the currently relevant payoff function.

Under these conditions, the "winning" perceptual strategies were (of course) still driven by their underlying fitness, and in this sense were still "interface" systems. However, the key result of our simulations (as depicted by everything in Fig. 3 above the red line for one task) was that when these two tweaks were made, the distinction between interface and veridical systems collapsed—with those interface systems that were *not* also veridical being driven to near extinction when faced with the demands of numerous tasks (holding out in only 3% of simulations when there were 2000 tasks).

These results—which effectively "save the day" for perceptual realism—are even more striking insofar as our simulations were designed to "stack the deck" *against* veridical strategies in at least four ways. *First*, all monotonic payoff functions were excluded, because they produce interface systems that trivially align with veridicality. *Second*, payoff functions in our primary simulations were *randomly* sampled—such that they were often very similar to each other. (Rather than purposefully selecting maximally distinct payoff functions, we randomly sampled payoff functions without biasing them toward being distinct.) But as we then showed in the secondary simulations, even sampling only *two* payoff functions with malice aforethought to be maximally distinct (minimally overlapping) is sufficient to give veridicality

the evolutionary edge. (And of course, in real-world contexts, it often seems trivially easy to recognize multiple tasks that seem as different as can be—e.g., when using a stick as kindling for a fire vs. a mast for a sail.) *Third*, because the space of possible nonveridical systems is much larger than the space of possible veridical systems (2024 nonveridical to 24 veridical), the initial population of (entirely randomly selected) strategies was overwhelmingly (∼99%) nonveridical. *Fourth*, by the same token, mutations were always overwhelmingly (by an order of magnitude) more likely to result in nonveridical strategies than veridical ones (even when the reproducer is a veridical parent). Despite these four factors, our evolutionary simulations nevertheless found that an increased number and variety of tasks both drove perceptual strategies to become veridical.

4.1. Generalizing to more complex environments

In line with the original simulations used to support ITP, our simulations were quite simple—including only one perceptual dimension. This raises the question of whether and how our results would generalize to more complex environments with multiple perceptual dimensions. Suppose, for example, that resources and payoffs vary along multiple (N) dimensions, rather than just one, and that the agents' perceptual system is able to perceive along multiple (M) dimensions as well. For some intuition, imagine that resources are determined by two features that are accessible to a perceptual system. Resources can then be represented as points in a 2D space, utility functions as functions over that 2D space, and the perceptual system as transforming each dimension into percepts. These two dimensions could be orthogonal to one another, but do not need to be. We can imagine the perceptual system mapping each dimension into two categories: red and green for dimension 1, and perhaps big and small for dimension 2. Over the course of evolution, the perceptual system will tend toward the optimal mapping for each dimension. So long as the number of distinct tasks (or 2D utility functions) far outnumbers the dimensions along which the agent can perceive, the perceptual system will be unable to successfully specialize for individual tasks without detriment to the other tasks. On the other hand, if the perceptual system can perceive along more dimensions than there are tasks, then the perceptual system could evolve to perceive mappings between each dimension and the utility function for a particular task. In other words, perceptual mappings for each dimension could specialize for just one or a few particular tasks, leading to an interface perceptual system. But so long as the number and variety of tasks overwhelm the dimensions that the perceptual system can perceive, this entails that our results will hold in a space of any dimension.

In a similar vein, we might also ask about whether and how our results depend on the types of payoff functions, and whether it matters that resources be ordered along a dimension. A key assumption underpinning our modeling (and the simulations supporting ITP) is that there is a relationship between the physical dimension along which the resources vary and the payoff, such that two resources near to each other along the physical dimension are also near to each other in payoff value. Were the values of each resource instead uniformly random and unrelated to the position of the resource along some physical dimension, perceiving that dimension would be utterly useless, and we would expect the distribution of perceptual

systems at each generation to be uniform over all possible perceptual strategies. Similarly, for multiple tasks where each payoff function is "jumpy" (e.g., a sinusoidal function with a high frequency and amplitude), such that two resources nearby on this dimension are mapped to very different payoffs, we would expect a uniform distribution of perceptual strategies. The assumption that similar resources have similar payoffs, almost by definition, must be valid for any perceptual system—since if there were no connection between a physical dimension and fitness payoffs, why perceive it in the first place? Furthermore, both the original ITP simulations and ours rely on the assumption that the stimuli/resources can be ordered along some dimensions. If the stimuli/resources cannot be ordered, then there can be no systematic connection between a physical dimension and fitness payoff. We are, therefore, comfortable limiting the generality of our results to situations where there is indeed a relationship between the ordering of the stimuli along a physical dimension and fitness payoffs.

4.2. Conclusion

Our evolutionary simulations suggest that the degree to which a perceptual system represents the world accurately may depend on the variety of tasks that it is used for—with more tasks driving greater veridicality. As such, it seems likely that humans perceive the world in a highly veridical way, given the vast number of goals that we can clearly entertain (from war and child-rearing to food and politics, and on and on). Of course, this rich and multidimensional "goalscape" may not apply to all organisms: perhaps a dung beetle, for example, has only a single goal for its ball of dung. And perhaps ITP is an appropriate description for organisms like beetles with a relatively limited goalscape and range of behavior. But for humans, the combined pressures of flexible goals and inflexible perceptual systems lead us to perceive the world as it is, after all.

Acknowledgments

For helpful conversations and/or comments on previous drafts, we thank Laurie Paul, Manish Singh, and several anonymous reviewers. We also thank Don Hoffman, Justin Mark, and Manish Singh for generously sharing the details of their original simulations. This project was funded by ONR MURI #N00014-16-1-2007 awarded to BJS, NSF grant BCS-2045778 awarded to JJE, and an NSF Graduate Research Fellowship awarded to RWT. Preliminary results were presented at the 2021 meeting of the *Vision Sciences Society*. The full data and code from our simulations are available at: https://osf.io/vhb87/?view_only=37b59aef778c48cba5a4efd65e4d72de

Notes

1 Following the original ITP proposals, we will refer to "veridicality" in this paper, in order to meet ITP on its own rhetorical terms. This previous work operationally defined "veridical perception" as a "homomorphism that preserves all structures" of a subset of

the objective world (with the "subset" clause excluding the perception of X-rays and quarks, etc.; Hoffman, Singh, & Prakash, 2015, p. 1484). We will similarly adopt this conception, while acknowledging that other notions of veridicality have also been developed in cognitive science and philosophy (e.g., based on "reliability"; Goldman, 1979; see also Burge, 2010), and that this work often requires considerable theoretical nuance. We will not explore such nuances in the current paper, since such details will not matter for our primary points or computational demonstrations (and since the original ITP work also does not do so).

2 In other recent work, Hoffman and colleagues have also analyzed such evolutionary questions using analytical proofs rather than evolutionary simulations (e.g., Prakash, 2020; Prakash, Fields, Hoffman, Prentner, & Singh, 2020, 2021). But this work still shares the same key assumptions—and problems—as do the computational projects. Here, we focus on the computational project since it will help to make the underlying challenges and implications of the interface theory especially salient—but the same points end up applying as well to the analytical arguments.

3 In the current context, note that cognitive impenetrability would not forbid different mappings from *percepts* to *actions* depending on the current goal. Cognitive impenetrability implies that the goal context cannot affect the mappings from *stimuli* to *percepts,* but says nothing about how those percepts are used downstream to guide actions. So, an agent with a cognitively impenetrable perceptual system can readily act differently when under the influence of different goals, even while perception itself remains impenetrable to goals.

4 In the context of these simulations, ITP assumes that any given perceptual function is equally costly to implement (e.g., in terms of time or energy)—and accordingly, our models also make that same assumption. We note in passing, though, that this may also effectively "stack the deck" in favor of ITP, since certain strategies are likely to be more difficult to implement in practice. In particular, those (veridical) strategies which need only make a binary distinction (between levels along the ground-truth dimension that are greater vs. less than some particular threshold, corresponding to the single "run" of a given color in such strategies) may be easier to implement than those (interface) strategies which need to distinguish multiple different levels (corresponding to multiple different "runs" of "r" or "g" in such strategies).

5 This mutation rate affects the speed of convergence and the ceiling (i.e., the asymptote for the highest proportion of the total population that any one strategy can achieve)—since if the mutation rate is very high, no one strategy can dominate.

6 Lower values of lambda produce payoff functions with a larger range of concentrations, or peakedness. The choice of lambda = 1/15 for the primary set of simulations was designed to produce this variety in the peakedness of the sampled utility functions. The goal in the secondary set of simulations, in contrast, was to explore the effect of overlap, and so sampling less concentrated payoff functions (using lambda = 1) allowed for larger areas of overlap that could be studied richly, which would not have been possible with extremely peaked distributions.

7 The qualitative pattern of results reported here does not depend on factors such as the exact number of possible resources, the exact number of colors, the exact number of individuals in the population, or the exact number of rounds in a generation.
8 Of this 8% of simulations, the median ratio of the more common color to the less common is 3.58, with a range from 1.75 (from a strategy like {4r, 7g}, with a ratio of 7:4) to 4.5 (from a strategy like {2r, 9g}, with a ratio of 9:2).
9 The perceptual strategies {5r, 6g}, {6r, 5g }, {6g, 5r}, and {5g, 6r} are treated here as variations of the same strategy: they all divide up the perceptual space roughly evenly, and it does not matter whether red or green comes first. There are information-theoretic reasons why dividing up the space evenly might be beneficial, but that is beyond the scope of our current discussion. (To briefly provide the intuition, learning that a resource is "red" or "green" will be most informative—or will reduce entropy the most—when the proportion of space that "red" and "green" span is equal.)

References

Anderson, B. L. (2015). Where does fitness fit in theories of perception? *Psychonomic Bulletin & Review*, *22*, 1507–1511.

Angelucci, A., Fano, V., Ferretti, G., Macrelli, R., & Tarozzi, G. (2021). Evolutionary dynamics and accurate perception. Critical realism as an empirically testable hypothesis. *Philosophia Scientiæ*, *25*, 157–178.

Berkeley, G. (1709). *An essay towards a new theory of vision*. Aaron Rhames.

Berkeley, G. (1725). *Three dialogues between Hylas and Philonous*. William & John Innys.

Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1076–1096.

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

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

Cohen, J. (2015). Perceptual representation, veridicality, and the interface theory of perception. *Psychonomic Bulletin & Review*, *22*, 1512–1518.

Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review*, *16*, 964–969.

Edelman, S. (2015). Varieties of perceptual truth and their possible evolutionary roots. *Psychonomic Bulletin & Review*, *22*, 1519–1522.

Feldman, J. (2015). Bayesian inference and "truth": A comment on Hoffman, Singh, and Prakash. *Psychonomic Bulletin & Review*, *22*, 1523–1525.

Fields, C. (2015). Reverse engineering the world: A commentary on Hoffman, Singh, and Prakash, "The interface theory of perception". *Psychonomic Bulletin & Review*, *22*, 1526–1529.

Firestone, C., & Scholl, B. J. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science*, *25*, 38–46.

Firestone, C., & Scholl, B. J. (2015a). Can you experience top-down effects on perception? The case of race categories and perceived lightness. *Psychonomic Bulletin & Review*, *22*, 694–700.

Firestone, C., & Scholl, B. J. (2015b). Enhanced visual awareness for morality and pajamas? Perception vs. memory in 'top-down' effects. *Cognition*, *136*, 409–416.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral & Brain Sciences*, *39*, 1–77.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Hillsdale, NJ: Erlbaum.

Gilchrist, A. (2020). The integrity of vision. *Perception*, *49*, 999–1004.

Goldman, A. I. (1979). What is justified belief? In G. S. Pappas (Ed.), *Justification and knowledge* (pp. 1–23). Dordrecht: Springer.
Hickok, G. (2015). The interface theory of perception: The future of the science of the mind? *Psychonomic Bulletin & Review*, *22*, 1477–1479.
Hoffman, D. D. (2009). The user-interface theory of perception: Natural selection drives true perception to swift extinction. In S. J. Dickinson, A. Leonardis, B. Schiele, & M. Tarr (Eds.), *Object categorization: Computer and human vision perspectives* (pp. 148–166). Cambridge University Press.
Hoffman, D. D. (2014). The origin of time in conscious agents. *Cosmology*, 18, 494–520.
Hoffman, D. D. (2016). The interface theory of perception. *Current Directions in Psychological Science*, *25*, 157–161.
Hoffman, D. D. (2018). The interface theory of perception. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (pp. 731–754). Wiley.
Hoffman, D. D. (2019). *The case against reality: Why evolution hid the truth from our eyes*. WW Norton & Company.
Hoffman, D. D., & Prakash, C. (2014). Objects of consciousness. *Frontiers in Psychology*, *5*, 577.
Hoffman, D. D., & Singh, M. (2012). Computational evolutionary perception. *Perception*, *41*, 1073–1091.
Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin & Review*, *22*, 1480–1506.
Jackson, F. (1977). *Perception: A representative theory*. Cambridge University Press.
Jansen, F. K. (2018). Hoffman's interface theory from a bio-psychological perspective. *NeuroQuantology*, *16*, 92–101.
Koenderink, J. (2015). Esse est percipi & verum factum est. *Psychonomic Bulletin & Review*, *22*, 1530–1534.
Korman, D. Z. (2019). Debunking arguments in metaethics and metaphysics. In A. Goldman & B. McLaughlin (Eds.), *Metaphysics and cognitive science* (pp. 377–363). Oxford University Press.
Kruschke, J. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
Locke, J. (1690). *An essay concerning human understanding*.
Mark, J. T. (2013). *Evolutionary pressures on perception: When does natural selection favor truth?* Irvine, CA: University of California.
Mark, J. T., Marion, B. B., & Hoffman, D. D. (2010). Natural selection and veridical perception. *Journal of Theoretical Biology*, *266*, 504–515.
Martínez, M. (2019). Usefulness drives representations to truth: A family of counterexamples to Hoffman's Interface Theory of Perception. *Grazer Philosophische Studien*, *96*, 319–341.
Mausfeld, R. (2015). Notions such as "truth" or "correspondence to the objective world" play no role in explanatory accounts of perception. *Psychonomic Bulletin & Review*, *22*, 1535–1540.
McKay, R. T., & Dennett, D. (2009). The evolution of misbelief. *Behavioral & Brain Sciences*, *32*, 493–510.
McLaughlin, B. P., & Green, E. J. (2015). Are icons sense data? *Psychonomic Bulletin & Review*, *22*, 1541–1545.
O'Connor, C. (2014). Evolving perceptual categories. *Philosophy of Science*, *81*, 840–851.
Pizlo, Z. (2015). Philosophizing cannot substitute for experimentation: Comment on Hoffman, Singh & Prakash (2014). *Psychonomic Bulletin & Review*, *22*, 1546–1547.
Popper, K. R. (1987). *Evolutionary epistemology, rationality, and the sociology of knowledge*. Open Court Publishing.
Prakash, C. (2020). On invention of structure in the world: Interfaces and conscious agents. *Foundations of Science*, *25*, 121–134.
Prakash, C., Fields, C., Hoffman, D. D., Prentner, R., & Singh, M. (2020). Fact, fiction, and fitness. *Entropy*, *22*, 514.
Prakash, C., Stephens, K. D., Hoffman, D. D., Singh, M., & Fields, C. (2021). Fitness beats truth in the evolution of perception. *Acta Biotheoretica*, *69*, 319–341.
Robertson, H. (1994). *Perception*. London: Routledge.
Schlesinger, M. (2015). The interface theory of perception leaves me hungry for more: Commentary on Hoffman, Singh, and Prakash, "The interface theory of perception". *Psychonomic Bulletin & Review*, *22*, 1548–1550.

Searle, J. (2015). *Seeing things as they are: A theory of perception*. New York: Oxford University Press.
Shepard, R. N. (1990). Possible evolutionary basis for trichromacy. *Proceedings of the SPIE/SPSE Symposium on Electronic Imaging: Science and Technology*, *1250*, 301–309.
Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 495–532). Oxford University Press.
Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review*, *1*, 2–28.
Shepard, R. N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral & Brain Sciences*, *24*, 581.
Thompson, E. (1995). Colour vision, evolution, and perceptual content. *Synthese*, *104*, 1–32.
Wilson, A. D. (2021). Interface theory vs Gibson: An ontological defense of the ecological approach. *Philosophical Psychology*, *36*, 989–1010.
Woods, A. J., Philbeck, J. W., & Danoff, J. V. (2009). The various perceptions of distance: An alternative view of how effort affects distance judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1104–1117.
van Buren, B., & Scholl, B. J. (2018). Visual illusions as a tool for dissociating seeing from thinking: A reply to Braddick (2018). *Perception*, *47*, 999–1001.