The Interface Theory of Perception:

Natural Selection Drives True Perception To Swift Extinction

Donald D. Hoffman

1
The Interface Theory of Perception

A goal of perception is to estimate true properties of the world. A goal of categorization is to classify its structure. Aeons of evolution have shaped our senses to this end. These three assumptions motivate much work on human perception. I here argue, on evolutionary grounds, that all three are false. Instead, our perceptions constitute a *species-specific user interface* that guides behavior in a niche. Just as the icons of a PC's interface hide the complexity of the computer, so our perceptions usefully hide the complexity of the world, and guide adaptive behavior. This *interface theory of perception* offers a framework, motivated by evolution, to guide research in object categorization. This framework informs a new class of evolutionary games, called *interface games,* in which pithy perceptions often drive true perceptions to extinction.

1.1 Introduction

The jewel beetle *Julodimorpha bakewelli* is category challenged [11, 12]. For the male of the species, spotting instances of the category *desirable female* is a pursuit of enduring interest and, to this end, he scours his environment for telltale signs of a female's shiny, dimpled, yellow-brown elytra (wing cases). Unfortunately for him, many males of the species *Homo sapiens,* who sojourn in his habitats within the Dongara area of Western Australia, are attracted by instances of the category *full beer bottle* but not by instances of the category *empty beer bottle,* and are therefore prone to toss their emptied "stubbies" unceremoniously from their cars. As it happens, stubbies are shiny, dimpled, and just the right shade of brown to trigger, in the poor beetle, a category error. Male beetles find stubbies irresistible. Forsaking all normal females, they swarm the stubbies, genitalia everted, and doggedly try to copulate despite repeated glassy rebuffs. Compounding misfortune, ants of the species *Iridomyrmex discors* capitalize on the beetles' category errors; the ants sequester themselves near stubbies, wait for befuddled beetles, and consume them, genitalia first, as they persist in their amorous advances.

Categories have consequences. Conflating beetle and bottle led male *J. bakewelli* into mating mistakes that nudged their species to the brink of extinction. Their perceptual categories worked well in their niche: Males have low parental investment and thus their fitness is boosted if their category *desirable mate* is more liberal than that of females (as predicted by the theory of sexual selection, e.g., [7, 39]). But when stubbies invaded their niche, a liberal category transformed stubbies into Sirens, 370 milliliter amazons with matchless allure.

The bamboozled *bakewelli* illustrate a central principle of perceptual categorization, the

Principle of Satisficing Categories: *Each perceptual category of an organism, to the extent that the category is shaped by natural selection, is a satisficing solution to adaptive problems.*

This principle is key to understanding the provenance and purpose of perceptual categories: They are satisficing solutions to problems such as feeding, mating, and predation that are faced by all organisms in all niches. However, these problems take different forms in different niches and therefore require a diverse array of specific solutions. Such solutions are satisficing in that (1) they are, in general, only *local* maxima of fitness and (2) the fitness function depends not just on one factor, but on numerous factors, including the costs of classification errors, the time and energy required to compute a category, and the specific properties of predators, prey and mates in a particular niche. Furthermore, (3) the solutions depend critically on what adaptive structures the organism already has: It can be less costly to co-opt an existing structure for a new purpose than to evolve *de novo* a structure that might better solve the problem. A backward retina, for instance, with photoreceptors hidden behind neurons and blood vessels, is not the "best" solution *simpliciter* to the problem of transducing light but, at a specific time in the phylogenetic path of *H. sapiens,* it might have been the best solution given the biological structures then available. Satisficing in these three senses is, on evolutionary grounds, central to perception and therefore central to theories of perceptual categorization.

According to this principle, a perceptual category is a satisficing solution to adaptive problems only "to the extent that the category is shaped by natural selection." This disclaimer might seem to eviscerate the whole principle, to reduce it to the assertion that perceptual categories are satisficing solutions, except when they're not.

The disclaimer must stand. The issue at stake is the debate in evolutionary theory over adaptationism: To what extent are organisms shaped by natural selection versus other evolutionary factors, such as genetic drift and simple accident? The claim that a specific category is adaptive is an empirical claim, and turns on the details of the case. Thus, this disclaimer does not eviscerate the principle; instead, it entails that, although one expects most categories to be profoundly shaped by natural selection, each specific case of purported shaping must be carefully justified in the normal scientific manner.

1.2 The Conventional View

Most vision experts do not accept the principle of satisficing categories, but instead, tacitly or explicitly, subscribe to a different principle, the

Principle of Faithful Depiction: *A primary goal of perception is to recover, or estimate, objective properties of the physical world. A primary goal of perceptual categorization is to recover, or estimate, the objective statistical structure of the physical world.*

For instance, Yuille and Bülthoff [44] describe the Bayesian approach to perception in terms of faithful depiction: "We define vision as perceptual inference, the estimation of scene properties from an image or sequence of images ... there is insufficient information in the image to uniquely determine the scene. The brain, or any artificial vision system, must make assumptions about the real world. These assumptions must be sufficiently powerful to ensure that vision is well-posed for those properties in the scene that the visual system needs to estimate." On their view, there is a physical world that has objective properties and statistical structure (objective in the sense that they exist unperceived). Perception uses Bayesian estimation, or suitable approximations, to reconstruct the properties and structure from sensory data. Terms such as *estimate, recover,* and *reconstruct,* which appear throughout the literature of computational vision, stem from commitment to the principle of faithful depiction.

Geisler and Diehl [8] endorse faithful depiction: "In general, it is true that much of human perception is veridical under natural conditions. However, this is generally the result of combining many probabilistic sources of information (optic flow, shading, shadows, texture gradients, binocular disparity, and so on). Bayesian ideal observer theory specifies how, in principle, to combine the different sources of information in an optimal manner in order to achieve an effectively deterministic outcome" (p. 397).

Lehar [24] endorses faithful depiction: "The perceptual modeling approach reveals the primary function of perception as that of generating a fully spatial virtual-reality replica of the external world in an internal representation." (p. 375).

Hoffman [15] endorsed faithful depiction, arguing that to understand perception we must ask, "First, why does the visual system need to organize and interpret the images formed on the retinas? Second, how does it remain true to the real world in the process? Third, what rules of inference does it follow?" (p. 154).

Noë and Regan [25] endorse a version of faithful depiction that is sensitive to issues of attention and embodied perception, proposing that "Perceivers are right to take themselves to have access to environmental detail and to learn that the environment is detailed" (p. 576) and that "the environmental detail is present, lodged, as it is, right there before individuals and that they therefore have access to that detail by the mere movement of their eyes or bodies" (p. 578).

Purves and Lotto [29] endorse a version of faithful depiction that is diachronic rather than synchronic, i.e., that includes an appropriate history of the world, contending that "what observers actually experience in response to any visual stimulus is its accumulated statistical meaning (i.e., what the stimulus has turned out to signify in the past) rather than the structure of the stimulus in the image plane or its actual source in the present" (p. 287).

Proponents of faithful depiction will, of course, grant that there are obvious limits. Unaided vision, for instance, sees electromagnetic radiation only through a chink between 400 and 700 nm, and it fails to be veridical for objects that are too large or too small. But these proponents maintain that, for middle-sized objects to which vision is adapted, our visual perceptions are in general veridical.

1.3 The Conventional Evolutionary Argument

Proponents of faithful depiction offer an evolutionary argument for their position, albeit an argument different than the one sketched above for the principle of satisficing categories. Their argument is spelled out, for instance, by Palmer [27](p. 6) in his textbook *Vision Science,* as follows: "Evolutionarily speaking, visual perception is useful only if it is reasonably accurate. ... Indeed, vision is useful precisely because it is so accurate. By and large, *what* *you see is what you get.* When this is true, we have what is called **veridical perception** ... perception that is consistent with the actual state of affairs in the environment. This is almost always the case with vision ..." [emphases his].

The error in this argument is fundamental: Natural selection optimizes fitness, not veridicality. The two are distinct and, indeed, can be at odds. In evolution, where the race is often to the swift, a quick and dirty category can easily trump one more complex and veridical. The jewel beetle's *desirable female* is a case in point. Such cases are ubiquitous in nature and central to understanding evolutionary competition between organisms. This competition is predicated, in large part, on exploiting the nonveridical perceptions of predators, prey and conspecifics, using techniques such as mimicry and camouflage.

Moreover, as noted by Trivers [40], there are reasons other than greater speed and less complexity for natural selection to spurn the veridical: "If deceit is fundamental to animal communication, then there must be strong selection to spot deception and this ought, in turn, to select for a degree of self-deception, rendering some facts and motives unconscious so as not to betray—by the subtle signs of self-knowledge—the deception being practiced. Thus, the conventional view that natural selection favors nervous systems which produce ever more accurate images of the world must be a very naïve view of mental evolution."

So the claim that "vision is useful precisely because it is so accurate" gets evolution wrong by conflating fitness and accuracy; they are not the same and, as we shall see with simulations and examples, they are not highly correlated. This conflation is not a peripheral error with trivial consequences: Fitness, not accuracy, is the *objective function* optimized by evolution. (This way of saying it doesn't mean that evolution *tries* to optimize anything. It just means that what matters in evolution is raising more kids, not seeing more truth.) Theories of perception based on optimizing the wrong function can't help but be radically misguided. Rethinking perception with the correct function leads to a theory strikingly different from the conventional. But first, we examine a vicious circle in the conventional theory.

1.4 Bayes' Circle

According to the conventional theory, a great way to estimate true properties of the world is via Bayes' theorem. If one's visual system receives some images, I, and one wishes to estimate the probabilities of various world properties, W, given these images, then one needs to compute the conditional probabilities $P(W|I)$. For instance, I might be a movie of some dots moving in two dimensions, and W might be various rigid and nonrigid interpretations of those dots moving in three dimensions. According to Bayes' theorem, one can compute

$$P(W|I) = P(I|W)P(W)/P(I).$$

$P(W)$ is the *prior probability.* According to the conventional theory, this prior models the assumptions that human vision makes about the world, e.g., that it has three spatial dimensions, one temporal dimension, and contains three-dimensional objects, many of which are rigid. $P(I|W)$ is the *likelihood.* According to the conventional theory, this likelihood models the assumptions that human vision makes about how the world maps to images; it's like a rendering function of a graphics engine, which maps a pre-specified three-dimensional world onto a two-dimensional image using techniques like ray tracing with Gaussian dispersion. $P(I)$ is just a scale factor to normalize the probabilities. $P(W|I)$ is the *posterior,* the estimate human vision computes about the properties of the world given the images I. So the posterior, which determines what we see, depends crucially on the quality of our priors and likelihoods.

How can we check if our priors and likelihoods are correct? According to the conventional theory, we can simply go out and measure the true priors and likelihoods in the world. Geisler & Diehl [8], for instance, tell us, "In these cases, the prior probability and likelihood distributions are based on measurements of physical and statistical properties of natural environments. For example, if the task in a given environment is to detect edible fruit in background foliage, then the prior probability and likelihood distributions are estimated by measuring representative spectral illumination functions for the environment and spectral reflectance functions for the fruits and foliage" (p. 380).

The conventional procedure, then, is to measure the true values in the world for the priors and likelihoods, and use these to compute, via Bayes, the desired posteriors. What the visual system ends up seeing is a function of these posteriors and its utility functions.

The problem with this conventional approach is that it entails a vicious circle, which we can call

Bayes' Circle: *We can only see the world through our posteriors. When we measure priors and likelihoods in the world, our measurements are necessarily filtered through our posteriors. Using our measurements of priors and likelihoods to justify our posteriors thus leads to a vicious circle.*

Suppose, for instance, that we build a robot with a vision system that computes shape from motion using a prior assumption that the world contains many rigid objects [41]. The system takes inputs from a video camera, does some initial processing to find two-dimensional features in the video images, and then uses an algorithm based on rigidity to compute three-dimensional shape. It seems to work well, but we decide to double-check that the prior assumption about rigid objects that we built into the system is in fact true of the world. So we send our robot out into the world to look around. To our relief, it comes back with the good news that it has indeed found numerous rigid objects. Of course it did; that's what we programmed it to do. If, based on the robot's good news, we conclude that our prior on rigid objects is justified, we've just been bagged by Bayes' Circle.

This example is a howler, but precisely the same mistake prompts the conventional claim that we can validate our priors by measuring properties of the objective world. The conventionalist can reply that the robot example fails because it ignores the possibility of cross checking results with other senses, other observers, and scientific instruments. But such a reply hides the same howler, because other senses, other observers, and scientific instruments all have built in priors. None is a filter-free window on an objective (i.e., observation independent) world. Consensus among them entails, at most, agreement among their priors; it entails nothing about properties or statistical structures of an objective world.

It is, of course, possible to pursue a Bayesian approach to perception without getting mired in Bayes' circle. Indeed, Bayesian approaches are among the most promising in the field. Conditional probabilities turn up everywhere in perception, because perception is often about determining what is the best description of the world, or the best action to take, *given* (i.e., conditioned on) the current state of the sensoria. Bayes is simply the right way to compute conditional probabilities using prior beliefs, and Bayesian decision theory, more generally, is a powerful way to model the utilities and actions of an organism in its computation of perceptual descriptions.

But it is possible to use the sophisticated tools of Bayesian decision theory, to fully appreciate the importance of utilities and the perception-action loop, and still to fall prey to Bayes' circle—to conclude, as quoted from Palmer above, that "Evolutionarily speaking, visual perception is useful only if it is reasonably accurate."

1.5 The Interface Theory of Perception

The conventional theory of perception gets evolution fundamentally wrong by conflating fitness and accuracy. This leads the conventional theory to the false claim that a primary goal of perception is faithful depiction of the world. A standard way to state this claim is the

Reconstruction Thesis: *Perception reconstructs certain properties and categories of the objective world.*

This claim is too strong. It must be weakened, on evolutionary grounds, to a less tendentious claim, the

Construction Thesis: *Perception constructs the properties and categories of an organism's perceptual world.*

The construction thesis is clearly much weaker than the reconstruction thesis. One can, for instance, obtain the reconstruction thesis by starting with the construction thesis and *adding* the claim that the organism's constructs are, at least in certain respects, roughly isomorphic to the properties or categories of the objective world, thus qualifying them to be deemed reconstructions.

But the range of possible relations between perceptual constructs and the objective world is infinite; isomorphism is just one relation out of this infinity and, on evolutionary grounds, an unlikely one. Thus the reconstruction thesis is a conceptual straightjacket that constrains us to think only of improbable isomorphisms, and impedes us from exploring the full range of possible relations between perception and the world. Once we dispense with the straightjacket we're free to explore all possible relations that are compatible with evolution [23].

To this end we note that, to the extent that perceptual properties and categories are satisficing solutions to adaptive problems, they admit a *functional* description. Admittedly, a conceivable, though unlikely, function of perception is faithful depiction of the world. That's the function favored by the reconstruction thesis of the conventionalist. But once we repair the conflation of fitness and accuracy, we can consider other perceptual functions with greater evolutionary plausibility. To do so properly requires a serious study of the functional role of perception in various evolutionary settings. Beetles falling for bottles is one instructive example; in the next section we consider a few more.

But here it's useful to introduce a model of perception that can help us study its function without relapse into conventionalism. The model is the

Interface Theory of Perception: *The perceptions of an organism are a user interface between that organism and the objective world* [16, 17, 20].

This theory addresses the natural question, "If our perceptions are not accurate, then what good are they?" The answer becomes obvious for user interfaces. The colour, for instance, of an icon on a computer screen does not estimate, or reconstruct, the true colour of the file that it represents in the computer. If an icon is, say, green, it would be ludicrous to conclude that this green must be an accurate reconstruction of the true colour of the file it represents. It would be equally ludicrous to conclude that, if the colour of the icon doesn't accurately reconstruct the true colour of the file, then the icon's colour is useless, or a blatant deception. This is simply a naïve misunderstanding of the point of a user interface. The conventionalist theory that our perceptions are reconstructions is, in precisely the same manner, equally naïve.

Colour is, of course, just one example among many: The shape of an icon doesn't reconstruct the true shape of the file; the position of an icon doesn't reconstruct the true position of the file in the computer. A user interface reconstructs nothing. Its predicates and the predicates required for a reconstruction can be entirely disjoint: Files, for instance, have no colour.

And yet a user interface is useful despite the fact that it's not a reconstruction. Indeed, it's useful *because* it's not a reconstruction. We pay good money for user interfaces because we don't want to deal with the overwhelming complexity of software and hardware in a PC. A user interface that slavishly reconstructed all the diodes, resistors, voltages and magnetic fields in the computer would probably not be a best seller. The user interface is there to facilitate our interactions with the computer by hiding its causal and structural complexity, and by displaying useful information in a format that is tailored to our specific projects, such as painting or writing.

Our perceptions are a species-specific user interface. Space, time, position and momentum are among the properties and categories of the interface of *H. sapiens* that, in all likelihood, resemble nothing in the objective world. Different species have different interfaces. And, due to the variation that is normal in evolution, there are differences in interfaces among humans. To the extent that our perceptions are satisficing solutions to evolutionary problems, our interfaces are designed to guide adaptive behavior in our niche; accuracy of reconstruction is irrelevant. To understand the properties and categories of our interface we must understand the evolutionary problems, both phylogenetic and ontogenetic, that it solves.

1.6 User Interfaces in Nature

The interface theory of perception predicts that (1) each species has its own interface (with some variations among conspecifics and some similarities across phylogenetically related species), (2) almost surely, no interface performs reconstructions, (3) each interface is tailored to guide adaptive behavior in the relevant niche, (4) much of the competition between and within species exploits strengths and limitations of interfaces, and (5) such competition can lead to arms races between interfaces that critically influence their adaptive evolution. In short, the theory predicts that interfaces are essential to understanding the evolution and competition of organisms; the reconstruction theory makes such understanding impossible. Evidence of interfaces should be ubiquitous in nature.

The jewel beetle is a case in point. Its perceptual category *desirable female* works well in its niche. However, its soft spot for stubbies reveals that its perceptions are not reconstructions. They are, instead, quick guides to adaptive behavior in a stubbie-free niche. The stubbie is a so-called *supernormal stimulus,* i.e., a stimulus that engages the interface and behavior of the organism more forcefully than the normal stimuli to which the organism has been adapted. The bottle is shiny, dimpled, and the right colour of brown. But what makes it a supernormal stimulus is apparently its supernormal size. If so, then, contrary to the reconstruction thesis, the jewel beetle's perceptual category *desirable female* does not incorporate a statistical estimate of the true sizes of the most fertile females. Instead its category satisfices with "bigger is better." In its niche this solution is fit enough. A stubbie, however, plunges it into an infinite loop.

Supernormal stimuli have been found for many species, and all such discoveries are evidence against the claim of the reconstruction theory that our perceptual categories estimate the statistical structure of the world; all are evidence for species-specific interfaces that are satisficing solutions to adaptive problems. Herring gulls (*Larus argentatus*) provide a famous example. Chicks peck a red spot near the tip of the lower mandible of an adult to prompt the adult to regurgitate food. Tinbergen and Perdeck [38] found that an artificial stimulus that is longer and thinner than a normal beak, and whose red spot is more salient than normal, serves as a supernormal stimulus for the chick's pecking behaviors. The colour of the artificial beak and head matter little. The chick's perceptual category *food bearer,* or perhaps *food-bearing parent,* is not a statistical estimate of the true properties of food-bearing parents, but a satisficing solution in which longer and thinner is better and in which greater salience of the red spot is better. Its interface employs simplified symbols that effectively guide behavior in its niche. Only when its niche is invaded by pesky ethologists is this simplification unmasked, and the chick sent seeking what can never satisfy.

Simplified does not mean simple. Every interface of every organism dramatically simplifies the complexity of the world, but not every interface is considered by *H. sapiens* to be simple. Selective sophistication in interfaces is the result, in part, of competition between organisms in which the strengths in the interface of one's nemesis or next meal are avoided and its weaknesses exploited. Dueling between interfaces hones them and the strategies used to exploit them. This is the genesis of mimicry and camouflage, and of complex strategies to defeat them.

A striking example, despite brains the size of a pinhead, are jumping spiders of the genus *Portia* [13]. *Portia* is araneophagic, preferring to dine on other spiders. Such dining can be dangerous; if the interface of the intended dinner detects *Portia,* dinner could be diner. So *Portia* has evolved countermeasures. Its hair and colouration mimic detritus found in webs and on the forest floor; its gait mimics the flickering of detritus—a stealth technology cleverly adapted to defeat the interfaces of predators and prey. If *Portia* happens on a dragline (a trail of silk) left by the jumping spider *Jacksonoides queenslandicus,* odors from the dragline prompt *Portia* to use its eight eyes to hunt for *J. queenslandicus.* But *J. queenslandicus* is well camouflaged and, if motionless, invisible to *Portia.* So *Portia* makes a quick vertical leap, tickling the visual motion detectors of *J. queenslandicus* and triggering it to orient to the motion. By the time *J. queenslandicus* has oriented, *Portia* is already down, motionless, and invisible to *J. queenslandicus;* but it has seen the movement of *J. queenslandicus.* Once the eyes of *J. queenslandicus* are safely turned away, *Portia* slowly stalks, leaps, and strikes with its fangs, delivering a paralyzing dose of venom. *Portia's* victory exploits strengths of its interface and weaknesses in that of *J. queenslandicus.*

Jewel beetles, herring gulls and jumping spiders illustrate the ubiquitous role in evolution of species-specific user interfaces. Perception is not reconstruction, it is construction of a niche-specific, problem-specific, fitness-enhancing interface, which the biologist Jakob von Uexküll [42, 43] called an *Umwelt* or "self-world" [34]. Perceptual categories are endogenous constructs of a subjective *Umwelt,* not exogenous mirrors of an objective world.

The conventionalist might object that these examples are self-refuting, since they require comparison between the perceptions of an organism and the objective reality that those perceptions get wrong. Only by knowing, for instance, the objective differences between beetle and bottle can we understand a perceptual flaw of *J. backewelli.* So the very examples adduced in support of the interface theory actually support the conclusion that perceptual reconstruction of the objective world in fact occurs, in contradiction to the predictions of that theory.

This objection is misguided. The examples discussed here, and all others that might be unearthed by *H. sapiens,* are necessarily filtered through the interface of *H. sapiens,* an interface whose properties and categories are adapted for fitness, not accuracy. What we observe in these examples is not, therefore, mismatches between perception and a reality to which *H. sapiens* has direct access. Instead, because the interface of *H. sapiens* differs from that of other species, *H. sapiens* can, in some cases, see flaws of others that they miss themselves. In other cases, we can safely assume, *H. sapiens* misses flaws of others due to flaws of its own. And, in yet other cases, flaws of *H. sapiens* might be obvious to other species.

The conventionalist might further object, saying, "If you think that the wild tiger over there is just a perceptual category of your interface, then why don't you go pet it? When it attacks, you'll find out it's more than an *Umwelt* category, it's an objective reality."

This objection is also misguided. I don't pet wild tigers for the same reason I don't carelessly drag a file icon to the trash bin. I don't take the icon literally, as though it resembles the real file. But I do take it seriously. My actions on the icon have repercussions for the file. Similarly, I don't take my tiger icon literally but I do take it seriously. Aeons of evolution of my interface have shaped it to the point where I had better take its icons seriously or risk harm. So the conventionalist objection fails because it conflates taking icons seriously and taking them literally.

This conventionalist argument is not new. Samuel Johnson famously raised it in 1763 when, in response to the idealism of Berkeley, he kicked a stone and exclaimed "I refute it *thus*" [4] (1, p. 134). Johnson thus conflated taking a stone seriously and taking it literally. Nevertheless Johnson's argument, one must admit, has strong psychological appeal despite the non sequitur, and it is natural to ask why. Perhaps the answer lies in the evolution of our interface. There was, naturally enough, selective pressure to take its icons *seriously*; those who didn't take their tiger icons seriously came to early harm. But were there selective pressures not to take its icons *literally*? Did reproductive advantages accrue to those of our Pleistocene ancestors who happened not to conflate the serious and the literal? Apparently not, given the widespread conflation of the two in the modern population of *H. sapiens.* Hence, the very evolutionary processes that endowed us with our interfaces might also have saddled us with the penchant to mistake their contents for objective reality. This mistake spawned sweeping commitments to a flat earth and a geocentric universe, and prompted the persecution of those who disagreed. Today it spawns reconstructionist theories of perception. Flat earth and geocentrism were difficult for *H. sapiens* to scrap; some unfortunates were tortured or burned in the process. Reconstructionism will, sans the torture, prove even more difficult to scrap; it's not just this or that percept that must be recognized as an icon, but rather perception itself that must be so recognized. The selection pressures on Pleistocene hunter-gatherers clearly didn't do the trick, but social pressures on modern *H. sapiens,* arising in the conduct of science, just might.

The conventionalist might object that death is a counterexample: it should be taken seriously *and* literally. It is not just shuffling of icons.

This objection is not misguided. In death, one's body icon ceases to function and, in due course, decays. The question this raises can be compared to the following: When a file icon is dragged to the trash and disappears from the screen, is the file itself destroyed, or is it still intact and just inaccessible to the user interface? Knowledge of the interface itself might not license a definitive answer. If not, then to answer the question one must add to the interface a theory of the objective world it hides. How this might proceed is the topic of the next section.

The conventionalist might persist, arguing that agreement between observers entails reconstruction and provides important reality checks on perception. This argument also fails. First, agreement between observers may only be apparent: It is straightforward to prove that two observers can be functionally identical and yet differ in their conscious perceptual experiences [18, 19]; reductive functionalism is false. Second, even if observers agree, this doesn't entail the reconstruction thesis. The observers might simply employ the same constructive (but not *re*constructive) perceptual processes. If two PC's have the same icons on their screens, this doesn't entail that the icons reconstruct their innards. Agreement provides subjective consistency checks—not objective reality checks—between observers.

1.7 Interface and World

The interface theory claims that perceptual properties and categories no more resemble the objective world than Windows icons resemble the diodes and resistors of a computer. The conventionalist might object that this makes the world unknowable and is, therefore, inimical to science.

This misses a fundamental point in the philosophy of science: Data never determine theories. This under-determination makes the construction of scientific theories a creative enterprise. The contents of our perceptual interfaces don't determine a true theory of the objective world, but this in no way precludes us from creating theories and testing their implications. One such theory, in fact the conventionalist's theory, is that the relation between interface and world is, on appropriately restricted domains, an isomorphism. This theory is, as we have discussed, improbable on evolutionary grounds and serves as an intellectual straightjacket, hindering the field from considering more plausible options.

What might those options be? That depends on which constraints one postulates between interface and world. Suppose, for instance, that one wants a minimal constraint that allows probabilities of interface events to be informative about probabilities of world events. Then, following standard probability theory, one would represent the world by a measurable space, i.e., by a pair (W, Σ_W), where W is a set and Σ_W is a σ-algebra of measurable events. One would represent the user interface by a measurable space (U, Σ_U), and the relation between interface and world by a measurable function $f\colon W \rightarrow U$. The function f could be many-to-one, and the features represented by W disjoint from those represented by U. The probabilities of events in the interface (U, Σ_U) would be *distributions* of the probabilities in the world (W, Σ_W), i.e., if the probability of events in the world is μ, then the probability of any interface event $A \in \Sigma_U$ is $\mu(f^{-1}(A))$. Using this terminology, the problem of Bayes' circle, scouted above, can be stated quite simply: It is conflating U with W, and assuming that $f\colon W \rightarrow U$ is approximately 1 to 1, when in fact it's probably infinite to 1. This mistake can be made even while using all the sophisticated tools of Bayesian decision theory and machine learning theory.

The measurable-space proposal could be weakened if, for instance, one wished to accommodate quantum systems with noncommuting observables. In this case the event structures would not be σ-algebras but instead σ-additive classes, which are closed under countable *disjoint* union rather than under countable union [10], and f would be measurable with respect to these classes. This would still allow probabilities of events in the interface to be distributions of probabilities of events in the world. It would explain why science succeeds in uncovering statistical laws governing events in space-time, even though these events, and space-time itself, in no way resemble objective reality.

This proposal could be weakened further. One could give up the measurability of f, thereby giving up any quantitative relation between probabilities in the interface and the world. The algebra or class structure of events in the interface would still reflect an isomorphic subalgebra or subclass structure of events in the world. This is a nontrivial constraint: Subset relations in the interface, for instance, would genuinely reflect subset relations of the corresponding events in the world.

Further consideration of the interface might prompt us, in some cases, to weaken the proposal even further. Multistable percepts, for instance, in which the percept switches while the stimulus remains unchanged, may force us to reconsider whether the relation between interface and world is even a *function*: Two or more states of the interface might be associated to a single state of the world.

These proposals all assume, of course, that mathematics, which has proved useful in studying the interface, will also prove useful in modeling the world. We shall see.

The discussion here is not intended, of course, to settle the issue of the relation between interface and world, but to sketch how investigation of the relation may proceed in the normal scientific fashion. This investigation is challenging because we see the world through our interface, and it can therefore be difficult to discern the limitations of that interface. We are naturally blind to our own blindness. The best remedy at hand for such blindness is the systematic interplay of theory and experiment that constitutes the scientific method.

The discussion here should, however, help place the interface theory of perception within the philosophical landscape. It is not classical relativism, which claims that there is no objective reality, only metaphor; it claims instead that there is an objective reality that can be explored in the normal scientific manner. It is not naïve realism, which claims that we directly see middle-sized objects; nor is it indirect realism, or representationalism, which says that we see sensory representations, or sense data, of real middle-sized objects, and do not directly see the objects themselves. It claims instead that the physicalist ontology underlying both naïve realism and indirect realism is almost surely false: A rock is an interface icon, not a constituent of objective reality. Although the interface theory is compatible with idealism, it is not idealism, because it proposes no specific model of objective reality, but leaves the nature of objective reality as an open scientific problem. It is not a scientific physicalism that rejects the objectivity of middle-sized objects in favor of the objectivity of atomic and subatomic particles; instead it claims that such particles, and the space-time they inhabit, are among the properties and categories of the interface of *H. sapiens.* Finally, it differs from the utilitarian theory of perception [5, 30, 31], which claims that vision uses a bag of tricks (rather than sophisticated general principles) to recover useful information about the physical world; interface theory (1) rejects the physicalist ontology of the utilitarian theory, (2) asserts instead that space and time, and all objects that reside within them, are properties or icons of our species-specific user interface, and therefore (3) rejects the claim of the utilitarian theory that vision recovers information about preexisting physical objects in space-time. It agrees, however, with the utilitarian theory that evolution is central to understanding perception.

A conventionalist might object, saying, "These proposals about the relation of interface and world are fine as theoretical possibilities. But, in the end, a rock is still a rock." In other words, all the intellectual arguments in the world won't make the physical world—always obstinate and always irrepressible—conveniently disappear. The interface theorist, no less than the physicalist, must take care not to stub a toe on a rock.

Indeed. But in the same sense a trash-can icon is still a trash-can icon. Any file whose icon stubs its frame on the trash can will suffer deletion. The trash can is, in this way, as obstinate and irrepressible as a rock. But both are simplifying icons. Both usefully hide a world that is far more complex. Space and time do the same.

The conventionalist might further object, saying, "The proposed dissimilarity between interface and world is contradicted by the user-interface example itself. The icons of a computer interface perhaps don't resemble the innards of a computer, but they *do* resemble real objects in the physical world. Moreover, when using a computer to manipulate 3D objects, as in computer aided design, the computer interface is most useful if its symbols really resemble the actual 3D objects to be manipulated."

Certainly. These arguments show that an interface can sometimes resemble what it represents. And that is no surprise at all. But user interfaces can also *not* resemble what they represent, and can be quite effective precisely because they don't resemble what they represent. So the real question is whether the user interface of *H. sapiens* does in fact resemble what it represents. Here, I claim, the smart money says No.

1.8 Future Research on Perceptual Categorization

So what? So what if perception is a user-interface construction, not an objective-world reconstruction? How will this affect concrete research on perceptual categorization?

Here are some possibilities. First, as discussed already, current attempts to verify priors are misguided. This doesn't mean we must abandon such attempts. It does mean that our attempts must be more sophisticated; at a minimum they must not founder on Bayes' circle.

But that is at a minimum. Real progress in understanding the relation between perception and the world requires careful theory building. The conventional theory that perception approximates the world is hopelessly simplistic. Once we reject this facile theory, once we recognize that our perceptions are to the world as a user interface is to a computer, we can begin serious work. We must postulate, and then try to justify and confirm, possible structures for the world and possible mappings between world and interface. Clinging to approximate isomorphisms is a natural, but thus far fruitless, response to this daunting task. It's now time to develop more plausible theories. Some elementary considerations toward this end were presented in the previous section.

Our efforts should be informed by relevant advances in modern physics. Experiments by Alain Aspect [1, 2], building on the work of Bell [3], persuade most physicists to reject *local realism,* viz., the doctrine that (1) distant objects cannot directly influence each other (*locality*) and (2) all objects have pre-existing values for all possible measurements, before any measurements are made (*realism*). Aspect's experiments demonstrate that distant objects, say two electrons, can be *entangled,* such that measurement of a property of one immediately affects the value of that property of the other. Such entanglement is not just an abstract possibility, it is an empirical fact now being exploited in quantum computation to give substantial improvements over classical computation [6, 21]. Our untutored categories of space, time and objects would lead us to expect that two electrons a billion light years apart are separate entities; in fact, because of entanglement, they are a single entity with a unity that transcends space and time. This is a puzzle for proponents of faithful depiction, but not for interface theory. Space, time and separate objects are useful fictions of our interface, not faithful depictions of objective reality.

Our theories of perceptual categorization must be informed by explicit dynamical models of perceptual evolution, models such as those studied in evolutionary game theory [14, 26, 33]. Our perceptual categories are shaped *inter alia* by factors such as predators, prey, sexual selection, distribution of resources, and social interactions. We won't understand categorization until we understand how categories emerge from dynamical systems in which these factors interact. There are promising leads. Geisler and Diehl [8] simulate interactions between simplified predators and prey, and show how these might shape the spectral sensitivities of both. Komarova, Jameson and Narens [22] show how colour categories can evolve from a minimal perceptual psychology of discrimination together with simple learning rules and simple constraints on social communication. Some researchers are exploring perceptual evolution in foraging contexts [9, 32, 35]. These papers are useful pointers to the kind of research required to construct theories of categorization that are evolutionarily plausible. As a concrete example of such research, consider the following class of evolutionary games.

1.9 Interface Games

In the simplest interface game, two animals compete over three territories. Each territory has a food value and a water value, each value ranging from, say, 0 to 100. The first animal to choose a territory obtains its food and water values; the second animal then chooses one of the remaining two territories, and obtains its food and water values. The animals can adopt one of two perceptual strategies. The *truth* interface strategy perceives the exact values of food and of water for each territory. Thus the total information that *truth* obtains is $I_T = 3$ [territories] × 2 [resources per territory] × $\log_2 101$ [bits per resource] ≈ 39.95 bits. The *simple* interface strategy perceives only one bit of information per territory: if the food value of a territory is greater than some fixed value (say 50), *simple* perceives that territory as green, otherwise *simple* perceives that territory as red. Thus the total information that *simple* obtains is $I_S = 3$ bits.

It costs energy to obtain perceptual information. Let the energy cost per bit be denoted by c_e. Since the *truth* strategy obtains I_T bits, the total energy cost to *truth* is $I_T c_e$, which is subtracted from the sum of food and water values that *truth* obtains from the territory it chooses. Similarly, the total energy cost to *simple* is $I_S c_e$.

It takes t units of time to obtain one bit of perceptual information. If $t > 0$, then *simple* acquires all of its perceptual information before *truth* does, allowing *simple* to be first to choose a territory.

Assuming, for simplicity, that the food and water values are independent, identically distributed random variables with, say, a uniform distribution on the integers from 0 to 100, we can compute a matrix of expected payoffs:

$$\begin{array}{rcc} & \underline{\text{Truth}} & \underline{\text{Simple}} \\ \text{Truth:} & a & b \\ \text{Simple:} & c & d \end{array}$$

Here a is the expected payoff to *truth* if it competes against *truth*, b is the expected payoff to *truth* if it competes against *simple*, c is the expected payoff to *simple* if it competes against *truth*, and d is the expected payoff to *simple* if it competes against *simple*.

As is standard in evolutionary game theory, we consider a population of *truth* and *simple* players and equate payoff with fitness. Let x_T denote the frequency of *truth* players and x_S the frequency of *simple* players; the population is thus $\vec{x} = (x_T, x_S)$. Then, assuming players meet at random, the expected payoffs for *truth* and *simple* are, respectively, $f_T(\vec{x}) = ax_T + bx_S$ and $f_S(\vec{x}) = cx_T + dx_S$. The selection dynamics is then $x'_T = x_T[f_T(\vec{x}) - F]; x'_S = x_S[f_S(\vec{x}) - F]$, where primes denote temporal derivatives and F is the average fitness, $F = x_T f_T(\vec{x}) + x_S f_S(\vec{x})$.

If $a > c$ and $b > d$, then *truth* drives *simple* to extinction. If $a < c$ and $b < d$ then *simple* drives *truth* to extinction. If $a > c$ and $b < d$, then *truth* and *simple* are bistable; which goes extinct depends on the initial frequencies, $\vec{x}(0)$, at time 0. If $a < c$ and $b > d$ then *truth* and *simple* stably coexist, with the *truth* frequency given by $(d-b)/(a-b-c+d)$. If $a = c$ and $b = d$, then selection does not change the frequencies of *truth* and *simple.*

The entries in the payoff matrix described above will vary, of course, with the correlation between food and water values, with the specific value of food that is used by *simple* as the boundary between green and red, and with the cost c_e per bit of information obtained.

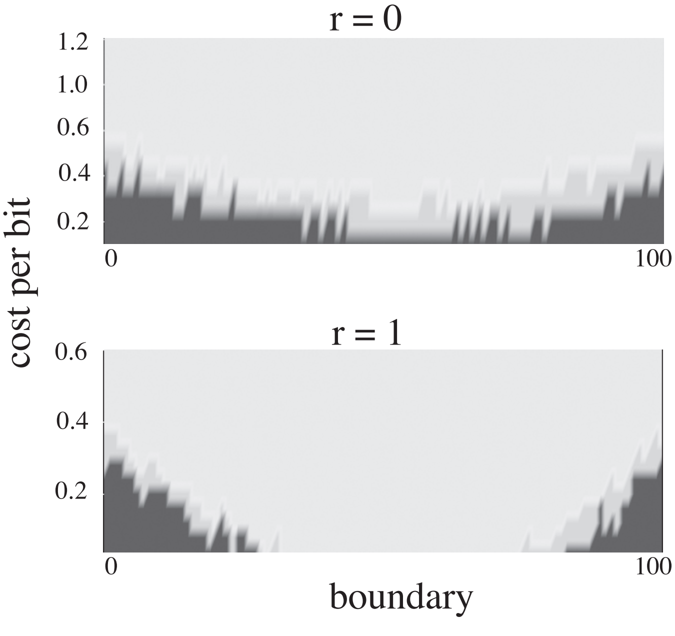

Fig 1.1. Asymptotic behavior of the interface game as a function of the cost per bit of information and the choice of the red-green boundary in the *simple* strategy. Light gray indicates that *simple* drives *truth* to extinction, intermediate gray that the two strategies coexist, and dark gray that *truth* drives *simple* to extinction. The

upper plot is for uncorrelated food and water, the lower for perfectly correlated food and water.

And here is the punchline. *Simple* drives *truth* to extinction for most values of the red-green boundary, even when the cost per bit of information is small and the correlation between food and water is small. This is illustrated in Figure 1.1, which shows the results of Matlab simulations. Evolutionary pressures do not select for veridical perception; instead they drive it, should it arise, to extinction.

The interface game just described might seem too simple to be useful. One can, however, expand on the simple game just described in several ways, including (1) increasing the number of territories at stake, (2) increasing the number of resources per territory, (3) having dangers as well as resources in the territories, (4) considering distributions other than uniform (e.g., Gaussian) for the resources and dangers, (5) considering two-boundary, three-boundary, n-boundary interface strategies, and more general categorization algorithms that don't rely on such boundaries, (6) considering populations with three or more interface strategies, (7) considering more sophisticated maps from resources to interfaces, including probabilistic maps, (8) considering time and energy costs that vary with architecture (e.g., serial versus parallel) and that are probabilistic functions of the amount of information gleaned and (9) extending the replicator dynamics, e.g., to include communication between players and to include a spatial dimension in which players only interact with nearby players (as has been done with stag hunt and Lewis signaling games [36, 37, 45]). Interface games, in all these varieties, allow us to explore the complex evolutionary pressures that shape perception and perceptual categorization, and to do so as realistically as our imaginations and computational resources will allow.

They will also allow us to address a natural question: As an organism's perceptions and behaviors become more complex, shouldn't it be the case that the goal of perception approaches that of recovering the properties of the environment?

Using simulations of interface games, one can ask for what environments (including what kinds of competitors) will the reproductive pressures push an organism to true perceptions of the environment, so that perceptual truth is an evolutionarily stable strategy. My bet: None of interest.

1.10 Conclusion

Most experts assume that perception estimates true properties of an objective world. They justify this assumption with an argument from evolution: Natural selection rewards true perceptions. I propose instead that if true perceptions crop up, then natural selection mows them down; natural selection fosters perceptions that act as simplified user interfaces, expediting adaptive behavior while shrouding the causal and structural complexity of the objective world. In support of this proposal, I discussed mimicry and mating errors in nature, and presented simulations of an evolutionary game.

Old habits die hard. I suspect that few experts will be persuaded by these arguments to adopt the interface theory of perception. Most will still harbor the long-standing conviction that, although we see reality through small portals, nevertheless what we see is, in general, veridical. To such experts I offer one final claim, and one final challenge. I claim that natural selection drives true perception to swift extinction: Nowhere in evolution, even among the most complex of organisms, will you find that natural selection drives truth to fixation, i.e., so that the predicates of perception (e.g., space, time, shape and color) approximate the predicates of the objective world (whatever they might be). Natural selection rewards fecundity, not factuality, so it shapes interfaces, not telescopes on truth [28] (p. 571). The challenge is clear: Provide a compelling counterexample to this claim.

Acknowledgements. Justin Mark collaborated in developing the interface games, and wrote the simulations presented in Figure 1.1. For helpful comments on previous drafts, I thank Ori Amir, Mike D'Zmura, Geoff Iverson, Carol Skrenes, Duncan Luce, Larry Maloney, Brian Marion, Justin Mark, Louis Narens, Steve Pinker, Kim Romney, John Serences, Brian Skyrms, and Joyce Wu. For helpful discussions I thank Mike Braunstein, Larry Maloney, Jon Merzel, Chetan Prakash, Rosie Sedghi, and Phat Vu.

References

1. Aspect, A., Grangier, P. and Roger, G. (1982a). Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment: A new violation of Bells inequalities. *Physical Review Letters* **49**, 91–94.
2. Aspect, A., Dalibard, J. and Roger, G. (1982b). Experimental test of Bells inequalities using time-varying analyzers. *Physical Review Letters* **49**, 1804–1807.
3. Bell. J.S. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics* **1**, 195–200.
4. Boswell, J. (1791). *The life of Samuel Johnson.*
5. Braunstein, M.L. (1983). Contrasts between human and machine vision: Should technology recapitulate phylogeny?, in *Human and machine vision,* ed. J. Beck, B. Hope, and A. Rosenfeld (Academic Press, New York).
6. Nielsen, M.A. and Chuang, I.L. (2000). *Quantum computation and quantum information.* (Cambridge University Press, Cambridge).
7. Daly, M. and Wilson, M. (1978). *Sex, evolution, and behavior.* (Duxbury Press, Massachusetts).
8. Geisler, W.S. and Diehl, R.L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science* **27**, 379–402.
9. Goldstone, R.L., Ashpole, B.C., and Roberts, M.E. (2005). Knowledge of resources and competitors in human foraging. *Psychonomic Bulletin & Review* **12**, 81–87.
10. Gudder, S. (1988). *Quantum probability.* (Academic Press, San Diego).
11. Gwynne, D.T. & Rentz, D.C.F. (1983). Beetles on the Bottle: Male Buprestids Make Stubbies for Females. *Journal of Australian Entomological Society* **22**, 79–80.
12. Gwynne, D.T. (2003). Mating mistakes, in *Encyclopedia of insects,* ed.V.H. Resh and R.T. Carde (Academic Press: San Diego).
13. Harland, D.P. & Jackson, R.R. (2004). Portia perceptions: The Umwelt of an Araneophagic jumping spider, in *Complex worlds from simpler nervous systems,* ed. F.R. Prete (MIT Press, Cambridge, MA).
14. Hofbauer, J. & Sigmund, K. *Evolutionary games and population dynamics.* (Cambridge University Press, Cambridge).
15. Hoffman, D. D. (1983). The interpretation of visual illusions. *Scientific American* **249**, 154–162.
16. Hoffman, D. D. (1998). *Visual intelligence: How we create what we see.* (W.W. Norton, New York).
17. Hoffman, D. D. (2006a). Mimesis and its perceptual reflections, in *A View in*

the Rear-Mirror: Romantic Aesthetics, Culture, and Science Seen from Today. Festschrift for Frederick Burwick on the Occasion of His Seventieth Birthday, ed. W. Pape (WVT, Wissenschaftlicher Verlag Trier: Trier) (Studien zur Englischen Romantik 3).

18. Hoffman, D.D. (2006b). The scrambling theorem: A simple proof of the logical possibility of spectrum inversion. *Consciousness and Cognition* **15**, 31–45.
19. Hoffman, D.D. (2006c). The scrambling theorem unscrambled: A response to commentaries. *Consciousness and Cognition* **15**, 51–53.
20. Hoffman, D.D. (2008, in press). Conscious realism and the mind-body problem. *Mind & Matter.*
21. Kaye, P., Laflamme, R. and Mosca, M. (2007). *An introduction to quantum computing.* (Oxford University Press: Oxford).
22. Komarova, N.L., Jameson, K.A. and Narens, L. (2007). Evolutionary models of color categorization based on discrimination. *Journal of Mathematical Psychology* **51**, 359–382.
23. Mausfeld, R. (2002). The physicalist trap in perception theory, in *Perception and the physical world,* ed. D. Heyer and R. Mausfeld (Wiley, New York).
24. Lehar, S. (2003). Gestalt isomorphism and the primacy of subjective conscious experience: A Gestalt Bubble model. *Behavioral and Brain Sciences* **26**, 375–444.
25. Noë, A, and Regan, J.K. (2002). On the brain-basis of visual consciousness: A sensorimotor account, in *Vision and mind: Selected readings in the philosophy of perception,* ed. A. Noë and E. Thompson (MIT Press, Cambridge, MA).
26. Nowak, M.A. (2006). *Evolutionary dynamics: Exploring the equations of life.* (Belknap/Harvard University Press, Cambridge, MA).
27. Palmer, S.E. (1999). *Vision science: Photons to phenomenology.* (MIT Press, Cambridge, MA).
28. Pinker, S. (1997). *How the mind works.* (W.W. Norton, New York).
29. Purves, D., and Lotto, R. B. (2003). *Why we see what we do: An empirical theory of vision.* (Sinauer, Sunderland, MA).
30. Ramachandran, V.S. (1985). The neurobiology of perception. *Perception* **14**, 97–103.
31. Ramachandran, V.S. (1990). Interactions between motion, depth, color and form: The utilitarian theory of perception, in *Vision: Coding and efficiency,* ed. C. Blakemore (Cambridge University Press, Cambridge).
32. Roberts, M.E. and Goldstone, R.L. (2006). EPICURE: Spatial and knowledge limitations in group foraging. *Adaptive Behavior* **14**, 291–313.
33. Samuelson, L. (1997). *Evolutionary games and equilibrium selection.* (MIT Press, Cambridge, MA).
34. Schiller, C.H. (1957). *Instinctive behavior: Development of a modern concept.* (Hallmark Press, New York).
35. Sernland, E., Olsson, O., and Holmgren, N.M.A. (2003). Does information sharing promote group foraging? *Proceedings of the Royal Society of London* **270**, 1137–1141.
36. Skyrms, B. (2002). Signals, evolution, and the explanatory power of transient information. *Philosophy of Science* **69**, 407–428.
37. Skyrms, B. (2004). *The stag hunt and the evolution of social structure.* (Cambridge University Press, Cambridge).
38. Tinbergen, N., and A. C. Perdeck. (1950). On the stimulus situation releasing the begging response in the newly hatched Herring Gull chick (Larus argentatus

argentatus Pont.). *Behaviour* **3**, 1–39.

39. Trivers, R.L. (1972). Parental investment and sexual selection, in *Sexual selection and the descent of man, 1871-1971,* ed. B. Campbell (Aldine Press, Chicago).
40. Trivers, R.L. (1976). Foreword, in R. Dawkins, *The selfish gene.* (Oxford University Press: New York).
41. Ullman, S. (1979). *The interpretation of visual motion.* (MIT Press, Cambridge, MA).
42. Von Uexküll, J. (1909). *Umwelt und Innenwelt der Tiere.* (Springer-Verlag, Berlin).
43. Von Uexküll, J. (1934). A stroll through the worlds of animals and men: A picture book of invisible worlds, reprinted in *Instinctive behavior: Development of a modern concept,* C.H. Schiller (1957) (Hallmark Press, New York).
44. Yuille, A., and Bülthoff, H. (1996). Bayesian decision theory and psychophysics, in *Perception as Bayesian inference,* ed. D. Knill and W. Richards (Cambridge University Press, Cambridge).
45. Zollman, K. (2005). Talking to neighbors: The evolution of regional meaning. *Philosophy of Science* **72**, 69–85.