Perception, 2012, volume 41, pages 1073 – 1091

doi:10.1068/p7275

Computational evolutionary perception

Donald D Hoffman[1], Manish Singh[2]
[1] Department of Cognitive Science, University of California, Irvine, CA 92697, USA;
e-mail: ddhoff@uci.edu; [2] Department of Psychology (and Center for Cognitive Science),
Rutgers University, New Brunswick, NJ 08901, USA; e-mail: manish@ruccs.rutgers.edu
Received 28 April 2012, in revised form 27 July 2012

Abstract. Marr proposed that human vision constructs "a true description of what is there". He argued that to understand human vision one must discover the features of the world it recovers and the constraints it uses in the process. Bayesian decision theory (BDT) is used in modern vision research as a probabilistic framework for understanding human vision along the lines laid out by Marr. Marr's contribution to vision research is substantial and justly influential. We propose, however, that evolution by natural selection does not, in general, favor perceptions that are true descriptions of the objective world. Instead, research with evolutionary games shows that perceptual systems tuned solely to fitness routinely outcompete those tuned to truth. Fitness functions depend not just on the true state of the world, but also on the organism, its state, and the type of action. Thus, fitness and truth are distinct. Natural selection depends only on expected fitness. It shapes perceptual systems to guide fitter behavior, not to estimate truth. To study perception in an evolutionary context, we introduce the framework of *Computational Evolutionary Perception* (CEP). We show that CEP subsumes BDT, and reinterprets BDT as evaluating expected fitness rather than estimating truth.

Keywords: evolution, fitness, natural selection, Bayesian decision theory, vision, evolutionary game theory

1 Human vision

What is vision, and what is it for? David Marr answered that "vision is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information". (Marr 1982, p. 31)

According to Marr, the descriptions produced by human vision make explicit the true shapes and positions of objects, along with such properties as their colors and textures. This is, Marr thought, "… the quintessential fact of human vision—that it tells about shape and space and spatial arrangement … it also tells about the illumination and about the reflectances of the surfaces that make the shapes … and about their motion". (p. 36)

The fact that vision is able to produce true descriptions of shapes and their spatial relations is, Marr proposed, a consequence of evolution by natural selection: "We … very definitely do compute explicit properties of the real visible surfaces out there, and one interesting aspect of the evolution of visual systems is the gradual movement toward the difficult task of representing progressively more objective aspects of the visual world". (p. 340)

Marr admitted that human vision falls prey now and then to illusion. A stick, for instance, protruding from water looks bent due to refraction of light by water. But Marr was confident that the occasional illusion does not contradict his claim that human vision produces, in general, true descriptions of the objective world: "… usually our perceptual processing does run correctly (it delivers a true description of what is there), but although evolution has seen to it that our processing allows for many changes (like inconstant illumination), the perturbation due to the refraction of light by water is not one of them". (p. 30)

This was no small point for Marr. His claim that human vision evolved to construct true descriptions of the physical world is, he thought, key to understanding how the scientific study of human vision must proceed: "… the visual system tries to deal only with physical things, using rules based on constraints supplied by the physical structure of the world to build up other descriptions that again have physical meaning. This means that extreme care is required in the formulation of theories because nature seems to have been very careful and exact in evolving our visual systems". (p. 75)

Why has natural selection shaped human vision to see objective properties of the world? Marr did not address this point at length, but gave this hint: "The payoff is more flexibility; the price, the complexity of the analysis and hence the time and size of brain required for it" (p. 340). Marr's idea is that we benefit by having a large repertoire of behaviors that allow us to act adaptively in many situations, but this repertoire needs visual descriptions that can guide, in each situation, the proper choice of behavior. Thus the utility of flexible behavior is, Marr proposed, a selection pressure that shaped human vision to produce, as he put it, "a true description of what is there".

The hypothesis that human vision constructs, for the most part, a true description of what is there is now, almost universally, accepted among vision scientists. For instance, Palmer in his textbook *Vision Science* argues that "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* … This is almost always the case with vision …" (Palmer 1999, p. 6). Knill and Richards (1996, p. 6) say: "Visual perception … involves the evolution of an organism's visual system to match the structure of the world and the coding scheme provided by nature". Noë and Regan (2002) propose 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). Geisler and Diehl (2002) say: "In general, (perceptual) estimates that are nearer the truth have greater utility than those that are wide of the mark" (p. 421).

2 Bayesian decision theory

The idea that perception is a process of inference that estimates descriptions of the world has a long intellectual history, going back at least to the *Optics* of the Islamic scholar Alhazen (965–1039; see Sabra 1978), through the work of Helmholtz (1910), and more recently of Gregory (1966, 1970, 1974) and Marr. The modern framework for modeling human vision along these lines is Bayesian decision theory (BDT; Geisler and Kersten 2002; Glimcher 2003; Jaynes 2003; Kersten et al 2004; Knill and Richards 1996; Maloney and Zhang 2010; Mamassian et al 2002). BDT provides a probabilistic analysis of vision at what Marr called the "computational level", ie an analysis of the formal problem that the visual system faces—the "outputs" it must compute given the retinal inputs, the constraints it brings to bear in doing so, etc—considered independently of the specific algorithm that it might use, and of details of the neural implementation.

The essential feature of problems in perception is that they are highly inductive in nature. Any image on the retina is consistent with a large number of possible scene interpretations. The only way the visual system can resolve the ambiguity is by bringing to bear additional constraints or "biases" based on its prior experience (phylogenetic and ontogenetic), and comparing the relative probabilities of different interpretations. Formally, the visual system is given an image (or set of images) i_0, and it must determine the "best" scene interpretation for that image. In probabilistic terms, it must compute the conditional, posterior, probability $p(s|i_0)$ for different possible scene interpretations s. To do so, it has available two sources of information. The first is the likelihood function, or the conditional probability $p(i_0|s)$, viewed as a function of s (recall that i_0 is fixed). The likelihood is high for those scene interpretations s that are consistent with the given image i_0—ie those scenes that could have generated the image i_0, and hence can "explain" it. Given the highly inductive nature of vision, however, there will typically be a large number of scene interpretations that are consistent with the image i_0. So the likelihood by itself will typically not suffice to resolve the ambiguity. The other source of information available to the visual system is the prior probability of different scenes $p(s)$. Certain scenes and states of the world are more likely than others—eg light is more likely to come from overhead than below—and such statistical regularities have presumably been "internalized" and embodied in visual processing (eg Feldman 2012; Geisler 2008; Jepson et al 1996; Shepard 1994).

Given the two probabilistic sources of information—the prior $p(s)$ and the likelihood $p(i_0|s)$—Bayes's formula provides a provably optimal way to combine them (Jaynes 2003):

$$p(s|i_0) = p(i_0|s)\, p(s)/p(i_0) \, .$$

In other words, the posterior is proportional to the product of the likelihood and the prior. In practice, the term $p(i_0)$ in the denominator plays no significant role when comparing the posterior for different scenes, because it involves no dependence on s.[1] Moreover, given that $p(s|i_0)$ is a probability distribution, it must integrate to 1 over all possible s; this constraint effectively turns $p(i_0)$ into a normalizing constant.

Application of Bayes's formula yields a probability distribution—the posterior distribution—on the space of scene interpretations. But in visual perception we don't see an entire distribution of possible visual scenes, we usually see just one scene. In order to pick a single "best" interpretation from this distribution, one must take into account the consequences of making errors. If, for instance, one is walking on a path by a cliff, misjudging distances on one side of the path can be fatal, but on the other side innocuous. The consequences of such errors are represented mathematically via the loss (or gain) function, which assigns for each possible error, or deviation from the "true" but unknown interpretation, a loss (or gain) value.[2] Different choices of loss functions lead to different ways of picking a single "best" scene interpretation from the posterior distribution (see, eg, Brainard and Freeman 1997; Geisler and Kersten 2002; Glimcher 2003; Mamassian et al 2002). One's final estimate is therefore a function of three factors: (i) the extent to which different interpretations can explain the image data (the likelihood); (ii) the "internal" biases or constraints the visual system has concerning various interpretations (the prior); and (iii) the penalties associated with deviations from the "true"—but unknown—interpretation (the loss function).

A BDT observer can certainly be prone to misperceptions or "illusions". For instance, if the assumptions embodied in its prior happen to be invalid in the present context (say, if it makes an assumption that light comes from overhead, but in the current scene light happens to come from below), then the estimate of the BDT observer will deviate systematically from the actual state of affairs. Thus BDT observers do not always see the truth. (Indeed, the above example makes it clear that they *could* not always see the truth: No matter what assumption a BDT observer might make, it would always be possible to place the observer in a context where that assumption is violated.) Hence, in this particular sense, BDT allows that one might not see the truth.

There is, however, a more fundamental sense in which BDT embodies the common assumption that vision has evolved to estimate the truth. Specifically, BDT assumes that the language of scenes—call it X—over which the posterior distribution is computed, does contain somewhere within it a true description of the world. One's estimate might miss this true description; but the truth is in X somewhere, even if the BDT observer happens to miss it

[1] When one compares two scene interpretations s_1 and s_2 by taking their posterior ratio $p(s_1|i_0)/p(s_2|i_0)$, the term $p(i_0)$ simply cancels out.

[2] Losses are simply negative gains.

in any given instance. It is in this way that the framework of BDT embodies the assumption that vision has evolved to see the truth. Our perceptual estimates are not always correct, but they are performed over a set of options that, somewhere within it, contains the truth.

This conceptualization is deeply linked to the historical roots of Bayesian methods—namely, as a means of computing "inverse probability". A context first considered by Laplace (1774) was that of computing the "probabilities of causes" from an observed event E. Here, one is interested in the probability $p(C_j|E)$ for each of various possible causes C_j, given the observation E. The probabilities one actually knows, however, are $p(E|C_j)$, namely, the conditional probability of observing the event E *if* the true cause were C_j. The problem is therefore one of "inverting" the conditional probabilities. Precisely the same applies to the BDT observer, of course, who knows $p(i_0|s)$ for different possible scenes, but whose goal is to compute $p(s|i_0)$. Since the mapping from 3D scenes to retinal images corresponds to "optics" (indeed, the likelihood function in vision applications is generally referred to as either the "projective mapping" or the "rendering function"), the goal of vision according to BDT is to invert the optical mapping. Vision therefore essentially becomes "inverse optics" (eg Adelson and Pentland 1996; Pizlo 2001).

An important implication of this "inverse probability" approach is that, in BDT, the space of perceptual interpretations is assumed to be identical with the space of objective world scenes (or world states). In the description above, the same set X played both roles—that of the objective world as well as the space of perceptual interpretations. This assumption is standard in BDT approaches to vision, and will become important for us later.

3 Primitive vision

Marr proposed that *human* vision constructs a true description of what is there. But he proposed a different, and less influential, account of *primitive*, or simpler, visual systems. He maintained that, for such systems, it is still the case that vision provides useful descriptions, but he thought that they accomplish this in a manner quite different from that of human vision.

He illustrated his idea with the house fly. Reichardt and Poggio (1980) found that images from the eyes of a fly are processed with a few distinct computations that provide just the information needed to control flight. One computation, for instance, analyzes the rate of visual-field expansion, and prompts the fly to land if this rate is large. Other computations analyze the angular direction and velocity of blobs of a certain angular size, and control the in-flight tracking of the fly. Each computation is fast, and focused on producing a description that is useful for some aspect of flight control.

Marr noted that these descriptions, in addition to being useful, are also not cluttered with irrelevant information: "… it is extremely unlikely that the fly has any explicit representation of the visual world around him—no true conception of a surface, for example, but just a few triggers and some specifically fly-centered parameters …" (p. 34). Marr cited other such examples: bug detectors in frog vision, hawk detectors in rabbit vision, and a detector of red Vs in the vision of jumping spiders that alerts them to the possible presence of a mate. In the discussion at the end of his book, Marr developed this point: "In a true sense, for example, the frog does not detect *flies*—it detects small, moving, black spots of about the right size. Similarly, the housefly does not really represent the visual world about it—it merely computes a couple of parameters … which it inserts into a fast torque generator and which cause it to chase its mate with sufficient frequent success". (p. 340)

Marr summarized these examples with the observation that "Visual systems like the fly's serve adequately and with speed and precision the needs of their owners, but they are not very complicated; very little objective information about the world is obtained. The information is all very subjective …". (p. 34)

Marr assumed that properties perceived by humans, such as shape and spatial relations, correspond to objective physical properties that exist independently of any observer. He proposed however that simpler visual systems, such as those of flies, frogs, and spiders, do not, in general, produce descriptions of *objective* properties of the world; instead they produce descriptions that are *subjective*, but useful (and sometimes complex; see Prete 2004). Thus, according to Marr, the visual world of a fly gives it little, if any, insight into the true nature of the physical world around it. It inhabits instead a world of subjective visual descriptions.

Why should this be? Marr answers that "One reason for this simplicity must be that these facts provide the fly with sufficient information for it to survive. Of course, the information is not optimal and from time to time the fly will fritter away its energy chasing a falling leaf or an elephant a long way away …" (p. 34). Marr's point is that vision evolves by natural selection. A visual process will pass from one generation to the next only if it confers enough fitness, ie "sufficient information for it to survive". Visual processes in the fly that construct subjective descriptions confer enough fitness to be retained by natural selection. As a result, the fly need see no objective truths to survive. Subjective descriptions suffice.

Marr offered another reason for subjective descriptions in the fly: "Another reason is certainly that translating these rather subjective measurements into more objective qualities involves much more computation" (p. 35). This is again an argument based on fitness: Objective description requires more computation, which usually requires more time and energy, and thus detracts from fitness, since computations that are faster and cheaper are, ceteris paribus, fitter. Extra calories spent on computation are extra calories that must be found and consumed. Extra milliseconds spent on computation are extra milliseconds that predators can use to attack, or prey can use to escape.

4 Subjective and yet useful

Marr claimed that perceptions of primitive visual systems might be subjective and yet useful. He contended that "… it is extremely unlikely that the fly has any explicit representation of the visual world around him—no true conception of a surface, for example, but just a few triggers and some specifically fly-centered parameters …", and yet he concluded that "visual systems like the fly's serve adequately and with speed and precision the needs of their owners …". It is natural to ask how this is possible. How can a visual system fail to see any objective properties of the world and yet succeed at serving its owner adequately? How is it possible to survive if one never sees the objective world?

The answer is that it is possible because *fitness* is not the same thing as the objective world, and evolution by natural selection favors perceptual strategies that, ceteris paribus, glean more information about fitness.

Fitness is distinct from the objective world because fitness depends not only on the objective world, but also on an organism, its current state, and type of action. Suppose, for instance, that the objective world contains a T-bone steak. For a hungry tiger looking to eat, that steak enhances fitness. For a sated tiger looking to mate it contributes no fitness. For a cow, in any state, it contributes no fitness. In this sense, fitness is not the same as the objective world.

Now consider two representational strategies for a visual system. A subjective strategy, that guides adaptive behavior but computes no objective descriptions of the world—such as Marr attributes to the fly—describes the expansion in the visual field; whereas an objective strategy describes the true shape of any nearby surface and the true time of arrival at that surface. Which description, subjective or objective, confers greater fitness? The answer, as we have just seen, depends on the organism, its state, and its action. For a fly in flight wanting to land, the subjective description is in fact computed. Thus we conclude that, in the niche in which flies evolved, the subjective description probably confers greater fitness than the objective.

Natural selection thus makes possible the evolution of an organism that is ignorant of the objective world, or, as Marr put it, that "does not really represent the visual world about it". For such an organism, subjective descriptions are fitter than objective. Evolution can fashion organisms that are ignorant of the objective world because natural selection depends only on fitness and not, *a priori*, on seeing truth. Marr proposed that, for flies, frogs, spiders, and many other organisms with simpler vision, ignorance of the objective world is not just possible, it is actual. They inhabit subjective visual worlds, ignorant of the objective world around them. And they survive—even thrive.

5 Perceptual strategies

Marr argued that evolution shaped human vision to construct "a true description of what is there", but shaped primitive vision, such as that of the fly, so that it "does not really represent the visual world about it". These are substantial claims about the evolution of perception.

They can be explored and tested using evolutionary game theory (eg Hofbauer and Sigmund 1998; Jameson and Komarova 2009; Maynard Smith 1982; Nowak 2006; Samuelson 1997; Sandholm 2007). An objective perceptual strategy that constructs "a true description of what is there" can be modeled precisely and allowed to compete against a subjective strategy that provides "just a few triggers and some specifically fly-centered parameters". One can study the fitness of these strategies in various worlds. One can ask: "Which worlds favor true perceptions, and which drive such perceptions to extinction?" Hunches about the effectiveness of perceptual strategies can be replaced with rigorous understanding.

To do this, we construct a mathematical framework that we call *computational evolutionary perception* (CEP). This framework defines and classifies perceptual strategies, and studies their evolution.

Let the set W represent the objective world and the set X represent certain perceptions of an organism. We know nothing a priori about W, but assume we can talk sensibly about probabilities of events in W, governed by a probability measure μ. (We note that W might have other structures on it, such as orders, groups, metrics, or vector spaces.)

A *perceptual strategy* is a random variable (more generally, a measurable function) $P: W \rightarrow X$ from the world W to the space of perceptions X. We will find it useful to think of perceptual strategies as communication channels between the objective world and the organism, that allow information to flow from the objective world to the organism.

There are four classes of perceptual strategies (Mark et al 2010). A *naive realist* strategy assumes that $X = W$, ie that one's perceptions correspond to and are exhaustive of the objective world, and requires that P preserves all structures in W, ie that the structures of one's perceptions correspond to the objective structures in the world. Intuitively, and in brief, a naive realist strategy sees the truth, the whole truth, and nothing but the truth. A *strong critical realist* strategy assumes $X \subset W$, ie that one's perceptions correspond to part but not necessarily all of the objective world, and requires that P projects all structures in W onto X, ie that the structures of one's perceptions correspond to the objective structures in the world. Intuitively, a strong critical realist sees part of the truth, but not necessarily the whole truth. A *weak critical realist* strategy allows that $X \not\subset W$, ie that one's perceptions need not correspond to any part of the objective world, and requires that P projects all structures in W onto X. Intuitively, a weak critical realist need not see the truth, but what it does see faithfully preserves relationships between elements of the truth. An *interface* strategy allows that $X \not\subset W$, ie that one's perceptions need not correspond to any part of the objective world, and does not require that P project all structures in W onto X. Intuitively, an interface strategy need not see the truth, and what it does see need not preserve relationships between elements of the truth (Hoffman 2009; Koenderink 2011). Nevertheless such perceptions can guide adaptive behaviors, just as the desktop of a computer guides adaptive behaviors of the user, while hiding the complexity of the hardware and software of the computer (Hoffman 2000).

When we say that weak critical realist and interface strategies "need not see the truth", we do not simply mean that these strategies can make mistakes, such as seeing a 3D shape that is not quite accurate or perceiving a stick in water as being bent. Instead, we mean that the *very language* in which these strategies express the possible range of perceptions may be the *wrong language* for describing any aspect of the objective world. For instance, to claim, in this sense, that when we perceive objects and surfaces we "need not see the truth", means that the objective world may ultimately contain no "objects" as we intuitively think of them. [Consider, for instance, that even in modern theories of physics, there is nothing that directly corresponds to our intuitive notion of objects and surfaces, such as tables and tabletops. Einstein (1936), in his *Physics and Reality* refers to the object concept as a "free creation of the human (or animal) mind" (p. 320).] In other words, these perceptual strategies allow for the possibility that, not just some detail here or there is off, but rather that the entire framework may be off.

It is critical that CEP includes the study of perceptual strategies that, in this more dramatic sense, need not see the truth. We want to ask if natural selection has shaped our perceptions to report true properties of the objective world, or whether, instead, it has shaped our perceptions to be entirely "subjective", as Marr claims is the case for the fly. In order to address this question rigorously, we must define and study perceptual strategies that result entirely in subjective perceptions, and ask how well they compete in evolutionary games against strategies that have objective perceptions.

To this end, CEP must canvas all possible perceptual strategies, which is why we just defined the four basic classes of strategies. These four classes of perceptual strategies form a nested hierarchy, the *Hierarchy of Perceptual Strategies*:

$$\textit{Naive realist} \subset \textit{Strong critical realist} \subset \textit{Weak critical realist} \subset \textit{Interface}\ .$$

That is, all naive realist strategies are strong critical realist strategies, but not vice versa; all strong critical realist strategies are weak critical realist strategies, but not vice versa; all weak critical realist strategies are interface strategies, but not vice versa. This is illustrated in figure 1.

Marr proposed that evolution shaped human vision to construct "a true description of what is there", but shaped simpler visual systems, such as that of the fly, so that it "does not really represent the visual world about it". His proposal can be restated in the language of this hierarchy: Human vision includes strong critical realist strategies, whereas the fly has interface strategies but no strong critical realist strategies.

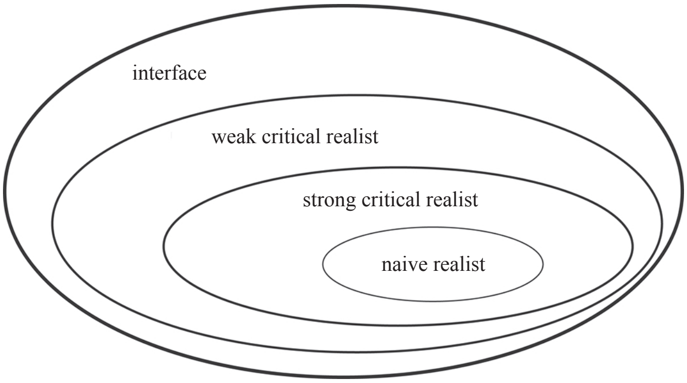

Figure 1. Venn diagram showing the inclusion relationship between the four classes of perceptual strategies.

6 Darwinian ideal observers

We wish to study proposals such as Marr's. To this end we must understand how natural selection shapes perceptual strategies. In competitions between strategies, selection favors the fitter strategy. Thus we require a notion of fitness for perceptual strategies.

We can think of organisms as gathering "fitness points" by interacting with the world. As mentioned earlier, fitness points depend on the organism, its state, and its action class. A hungry lion looking to eat would gain many fitness points from a gazelle, but a sated lion looking to mate would not. Natural selection favors perceptual strategies that guide an organism toward actions that reap more fitness points.

We formalize this by defining a *global fitness* function $f: W \times O \times S \times A \rightarrow \mathbb{R}$, where W denotes the unknown objective world, O a set of organisms, S a set of possible states of organisms, and A a set of action classes. Once we specify a fixed organism $o_i \in O$, state $s_j \in S$, and action class $a_k \in A$, then the fitness function f_{o_i, s_j, a_k} describes the fitness points that will be obtained for each $w \in W$. For instance, if the organism o_i is a lion, its state s_j is hunger, the action class a_k is eating, then the fitness function f_{o_i, s_j, a_k} describes the fitness points that will accrue from eating particular things (say, a gazelle). We call f_{o_i, s_j, a_k} a *specific fitness* function. Specific fitness functions guide evolution by natural selection.

A *Darwinian ideal observer* $\langle X, P \rangle$ for a specific fitness function f_{o_i, s_j, a_k} consists of (i) a set of perceptual representations X; and (ii) a perceptual strategy, or channel, P from the world W to X, ie $P: W \rightarrow X$, that allows the organism to maximize its channel capacity for expected fitness.[3] Such an observer is called *ideal* because natural selection does not, in general, produce perceptual strategies that *maximize* capacity for expected fitness. Satisficing solutions, not optimizing solutions, are the norm for natural selection. A perceptual strategy that has been shaped by natural selection as a satisficing solution for a specific fitness function f_{o_i, s_j, a_k} and set of perceptual representations X is a *Darwinian observer*.

To illustrate a Darwinian observer, consider a simple world with different territories, each territory containing just one resource that varies in quantity from 0 to 100, with each quantity having equal probability. In this simple setup, each value between 0 and 100 corresponds to a different "world state". Suppose that the specific fitness function guiding natural selection is bimodal, as illustrated in figure 2a.

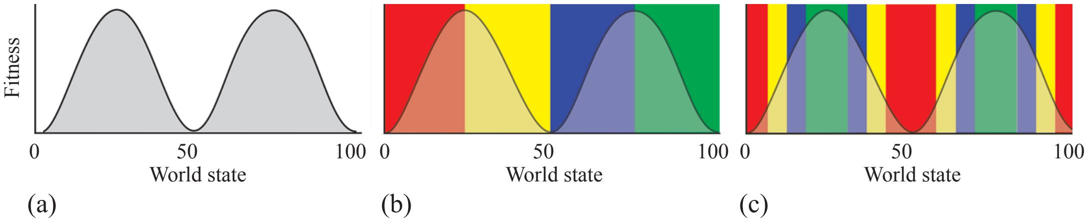

Figure 2. [In color online, see http://dx.doi.org/p7275] (a) A specific fitness function on a world having one resource that varies in quantity from 0 to 100. Worlds with resource quantities 25 or 75 confer highest fitness, and those with quantities 0, 50, or 100 confer least fitness. Fitness varies smoothly between these values according to the bell curves shown in the figure. (b) A perceptual strategy that estimates truth. High resource states are seen as *green*, low resource states as *red*, and intermediate resource states as either *blue* or *yellow*. (c) A Darwinian observer. High fitness states are seen as *green*, low fitness states as *red*, and intermediate fitness states as either *blue* or *yellow*.

[3] The capacity of a channel corresponds to the amount of information that can be transmitted over it. So a high-capacity channel for expected fitness means that the output of the channel—the representation in X—is (on average) highly informative about the fitness associated with different elements of W.

Suppose the organism has four possible percepts—say *red*, *yellow*, *blue*, and *green* (or R, Y, B, G, for short). In this case a perceptual strategy is a function from the interval $W = [0, 100]$ to the set $X = \{\mathrm{R, Y, B, G}\}$.

A critical-realist strategy that is a Bayesian ideal observer for estimating truth is illustrated in figure 2b. The world states [0, 25) are mapped to R, the states [25, 50) to Y, the states [50, 75) to B, and the states [75, 100] to G. Then the best formal estimate for the percept R is 12.5, for Y is 37.5, for B is 62.5, and for G is 87.5.[4] If we assume that the perceptions are ordered G > B > Y > R, then this perceptual strategy is critical realist because it preserves the order relation on the world states, ie all states that are seen as *green* are greater in quantity than all states that are seen as *blue*, and so on.

Although this perceptual strategy is ideal for estimating the true state of the world, it is useless for guiding adaptive behavior: The expected fitness of all states that are seen as *red* is identical to the expected fitness of all states seen as *yellow*, *blue*, or *green*. To see why this is so, look again at figure 2b. The resource states between 0 and 25 are seen as red. Over these states, the bell-shaped fitness curve goes from 0 fitness at resource state 0 to maximum fitness at resource state 25. The expected fitness for the states seen as red is therefore the total area under the bell curve from 0 to 25, divided by 25. Similarly, the expected fitness of the states seen as yellow, ie the states from 25 to 50, is the total area under the bell curve from 25 to 50, again divided by 25. But because of the symmetry of the bell curve, the area under the curve for states seen as yellow is identical to the area under the curve for states seen as red. Thus the expected fitness for the states seen as yellow is identical to the expected fitness of the states seen as red. Similar arguments hold for the expected fitness of states seen as blue and states seen as green.

Suppose, then, that the organism with the perceptual strategy of figure 2b is given a choice between two territories, one that has a resource quantity of 48—and therefore, as can be seen from the bell curves in figure 2b, has low fitness—and one that has a resource quantity of 71—and therefore, according to the same bell curve, has high fitness. The organism will see the first territory as *yellow* and the second as *blue*. Since the expected fitness of *yellow* is identical to that of *blue*, the organism must choose randomly between them. Its perceptions, although ideal for estimating truth, are useless for guiding adaptive behavior. Natural selection would deal harshly with such a perceptual strategy. In the language of information and communication theory (Cover and Thomas 2006), this Bayesian ideal observer is a communication channel with high channel capacity for messages about truth, but low capacity for messages about fitness.

An interface perceptual strategy that is a Darwinian observer is illustrated in figure 2c. World states with highest fitness (ie where the bell curves in the figure are highest) are mapped to *green*, states with moderately high fitness to *blue*, states with moderately low fitness to *yellow*, and states with lowest fitness to *red*. If the organism is given the same choice between two territories, one that has a resource quantity of 48, and one that has a resource quantity of 71, it will see the first territory as *red* and the second as *green*. Since the expected fitness of *green* is greater than that of *red*, the organism can successfully choose the territory offering greater fitness.

Although this strategy is effective for guiding adaptive choices, it is useless for estimating the true state of the world. If the organism sees *red*, then the corresponding world state could be near 0, near 50, or near 100. If it sees *yellow*, then the state could be near 10, 40, 60, or 90. For a loss function that is quadratic, the Bayesian estimate of the world state is exactly 50, no matter which color the organism sees. Thus this Darwinian observer is useless for discriminating the true state of the world. In the language of information theory, this Darwinian observer is a communication channel with high capacity for messages about fitness but low capacity for messages about truth.

[4] This assumes a so-called quadratic loss function.

The Darwinian observer and the Bayesian ideal observer each have the same set of perceptions: *red*, *yellow*, *blue*, *green*. Moreover, the perceptions of each observer are a function of the world state, ie of the number of resources in the world. But the functions are different. For the Bayesian ideal observer this function makes the perceptions informative about the true number of resources in the world, but not informative about the fitness of those resources. For the Darwinian observer this function makes the perceptions informative about the fitness of the resources, but not about the true number of resources. The perceptions of the Darwinian observer can help it to glean more fitness points, whereas the perceptions of the Bayesian ideal observer cannot. The Darwinian observer looks at the world through a lens that reveals fitness, the Bayesian ideal observer through a lens that reveals truth. Natural selection favors a lens that reveals fitness.

For ease of exposition, this example uses a discrete world (and discrete perceptions). However, nothing essential depends on the discreteness, and the same conclusion follows with continuous spaces. The example can also be readily extended to noisy perceptions (where the mapping from W to X is noisy) with the same result that the Darwinian observer gleans more fitness than the Bayesian ideal observer, and therefore is favored by natural selection.

In these examples, the Bayesian ideal observer is a critical-realist strategy whereas the Darwinian observer is an interface strategy that is not critical realist (a *strict interface* strategy). One naturally wonders whether this is generally the case. That is, a central question for computational evolutionary perception is this:

> Where, within the hierarchy of perceptual strategies, do Darwinian observers typically reside?

One would expect that most Darwinian observers—even Darwinian ideal observers—are strict interface strategies. The argument for this is as follows. Unless there is a very special relation between the structures of the world and those of the specific fitness function, optimizing a perceptual strategy for fitness will yield a solution no worse than that obtained by optimizing for fitness *subject to the constraint* that the perceptual strategy must project all structures of the world onto the set of perceptions (ie be a homomorphism). The only structure that a strict interface strategy must respect is the structure of probabilistic events; this is necessary for a perceptual strategy to be a random variable (or a measurable function). Requiring a perceptual strategy to respect additional structures should in general yield poorer solutions. Research by Mark et al (2010), using evolutionary game theory, finds that Darwinian ideal observers that are strict interface strategies outcompete naive-realist and critical-realist strategies in a variety of worlds. However, the general question must still be considered open, and further research is required.

7 Darwin and Bayes

Bayesian decision theory and Bayesian ideal observers are widely used in the study of human vision. Their application to vision typically assumes that human vision estimates the true state of the objective world or, as Marr put it, “a true description of what is there”. How is standard Bayesian decision theory (BDT) related to the framework of computational evolutionary perception (CEP) described here? In brief, CEP subsumes BDT and reinterprets it in evolutionary terms. Less briefly, their relationship is shown in the commuting diagram of figure 3.[5]

[5] To say that the diagram is *commuting* simply means that whenever there are two (or more) ways of getting from one place to another in the diagram, the two routes yield identical results.

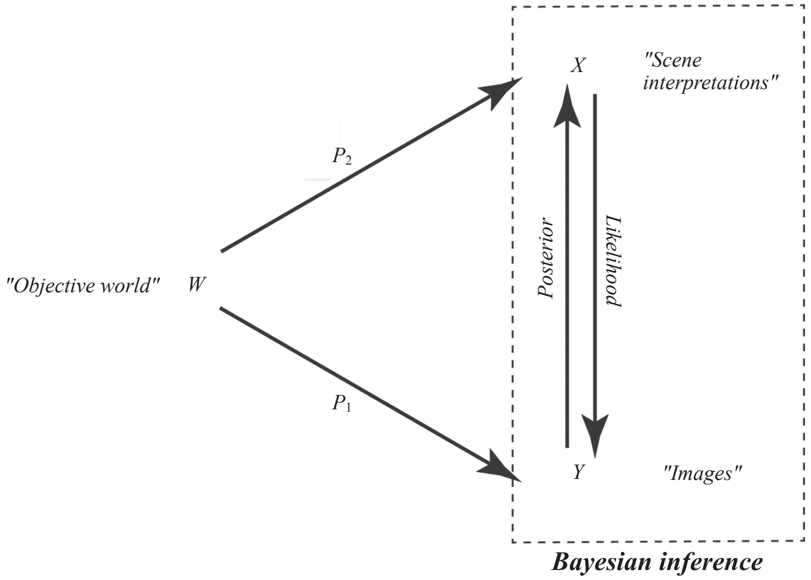

Figure 3. A commuting diagram showing the relationship between computational evolutionary perception (CEP) and Bayesian decision theory (BDT). The BDT framework corresponds to the portion of the diagram that lies within the boxed region on the right. By contrast, CEP does not assume that the interpretation space *X* corresponds to the objective world *W*. Hence in the CEP framework, the objective world *W* lies outside of the Bayesian inferential apparatus (for mathematical details see Singh and Hoffmann 2012).

On the left, *W* denotes the unknown objective world. The set *Y* denotes one set of perceptual representations, eg some of the image representations that Marr incorporated in his primal sketch. The set *X* denotes another set of perceptual representations, eg some representations of three-dimensional scenes.

The perceptual strategy $P_1 : W \rightarrow X$ is a Darwinian observer for the perceptual space *Y*. The perceptual strategy $P_2 : W \rightarrow X$ is a Darwinian observer for the perceptual space *X*.

The framework of BDT, as applied to perception, is the portion of figure 3 within the dashed box. The set *X* is interpreted within BDT as possible states of the external world. However in CEP the set *X* is *not* assumed to represent possible states of the objective external world; that role is played by *W*. CEP simply assumes that *X* is a set of perceptual representations that has been shaped by natural selection to guide adaptive behaviors. The representations in *X* need not resemble true states of the objective world *W*. Instead they are simply the outputs of a Darwinian observer

The set *Y* is interpreted within BDT as the possible perceptual inputs from which estimates in *X* are to be computed. In CEP, *Y* is simply a set of perceptual representations that are the outputs of a particular Darwinian observer. It might be that *Y* evolved before *X* (eg it is likely that representations of image structure, such as Marr's primal sketch, evolved before surface-based representations, such as Marr's 2½-D sketch), and that *X* evolved because Darwinian observers with the perceptual representations in *X* had substantially greater fitness than those that had only *Y*. This does not necessarily mean that *X* is closer to the truth *W* than is *Y*, only that *X* allows greater fitness.

Within BDT, the map from *X* to *Y* is the likelihood function, and the map from *Y* to *X* specifies the posterior distribution for any given element of *Y*.

The loss/gain function in BDT is a function that assumes there is an unknown true state $x_T \in X$ of the objective world, and it assigns losses or gains to various interpretations $x \in X$ as a function of their deviation ("error") from x_T (eg a Dirac function for MAP estimates, or quadratic function for mean estimates). The loss/gain function of BDT is *not* a fitness function that guides the evolution of perception by natural selection. That role is played by the specific fitness functions f_{o_i, s_j, a_k}. Confusion on this point has led some vision researchers to conclude that perceptions that deviate from the truth necessarily have less evolutionary utility.

The perceptual strategies $P_1 : W \rightarrow X$ and $P_2 : W \rightarrow X$ can be thought of as communication channels between the objective world, W, and perceptual representations of the organism. Natural selection tunes these channels to the only signal that matters for evolution, namely expected fitness. This tuning can be thought of as hill-climbing towards increasing channel capacity for expected fitness.[6]

8 Using the CEP framework to model perceptual evolution

A natural question is what can one do with the CEP framework that would not be possible with—or at least very difficult to accommodate within—BDT? As we noted earlier, consistent with the "inverse probability" conceptualization, BDT models of vision typically assume that the space of perceptual interpretations is identical with the objective world; and hence that the language of perceptual interpretations is the correct language for describing objective reality. Hence, BDT may be viewed as a special case of CEP in which $X = W$ (see figure 3).

In other words, CEP subsumes BDT because it allows for the possibility that the space of perceptual representations X that has been evolved by natural selection may not correspond to the objective truth. It is relatively easy to appreciate this in the context of primitive visual systems (such as the house fly we considered earlier). Perceptual representations can confer a great deal of fitness—allowing organisms to not only survive but thrive in their respective ecological niches—and yet have little to do with objective truth. Our framework allows for the possibility that the same may be true of human vision as well.

Below, we pose a number of questions on the evolution of perceptual systems that can be articulated quite naturally within the CEP framework, but would be difficult, perhaps even impossible, to capture within a standard BDT framework.

8.1 *Evolution of perceptual strategies or channels*

In section 6, we considered a simple example where the space of perceptual representations $X = \{\mathrm{R, Y, B, G}\}$ was fixed, and the fitness function was fixed (shown in figure 2a). We then considered two different perceptual strategies or channels—the *truth* strategy and the *Darwinian* strategy—illustrated in figures 2b and 2c, respectively. In the context of this example, the first strategy had essentially zero capacity for the fitness signal, whereas the second had high capacity.

This example motivates the following general question: Given a fixed space of perceptual representations Y, and a specific fitness function f, how can selection pressures shape the perceptual strategy, ie the channel P, to maximize its capacity—or at least hill-climb towards greater capacity—for the expected fitness signal? (Note that this question would not be meaningful if one were to assume that the mapping from W to Y is fixed, as is the likelihood function in BDT.)

8.2 *Evolution of perceptual representations*

In the evolution of biological perceptual systems, it is clear that perceptual representations themselves evolve (not just perceptual strategies for fixed representations). Whereas the visual systems of "lower" species compute relatively simple representations of image structure, advanced visual systems tend to compute more complex representations as well, eg involving some three-dimensional structure. Although it is fairly standard to take this as evidence that

[6] "Increasing channel capacity for expected fitness" means increasing the mutual information between the expected fitness signal and the output messages of the perceptual channel. In other words, the representation in X on average carries more and more information about the fitness associated with elements of W.

the more advanced visual systems are better tuned to objective world structure, this does not actually follow. What drives evolution is fitness, not truth, and selection pressures depend in potentially complicated ways on the *combination* of world structure and the fitness function f.

In the commuting diagram in figure 3, we considered two perceptual representations X and Y, such that Y—the more "primitive" representation—evolved first, and selection pressures subsequently pushed in the direction of evolving representation X. How can such selection pressures to evolve more complex perceptual representations be modeled? Note that this is qualitatively different from a situation in which the value of a parameter within a fixed representational format (say, the peak of a spectral sensitivity function) is nudged by evolutionary pressures. The question we are posing here concerns more dramatic changes that alter the very structure or format of a perceptual representation.

Within the framework of CEP, we can articulate the question precisely as follows: Given an objective world W, and a specific fitness function (ie for a fixed organism, state, and action class), under what conditions will a Darwinian observer[7] $\langle Y, P_1 \rangle$ evolve by altering the space of perceptual representations Y itself (rather than simply tuning the perceptual channel P_1 to the fixed space Y) to increase the channel's capacity for expected fitness? The key factor as always is expected fitness: Going to a different representational space X (with its own perceptual strategy or channel P_2) should result in a substantial increase in the (new) channel's capacity for expected fitness, relative to, say, simply tuning P_1 to increase its capacity given the fixed representational space Y.

Of course, X and Y are just two out of many possible perceptual representational spaces. In studying the evolution of perceptual representations, one must therefore consider evolutionary sequences of representational spaces $X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \ldots$. A sequence such as this can of course branch out: Two or more perceptual representations may evolve in parallel from a single "primitive" one—perhaps emphasizing different properties of the more primitive representation (such as shape, color, and motion). But one can nevertheless consider linear sequences obtained by following any ascending path through the lattice of evolving representational spaces.

One may then ask: Are there conditions under which such sequences would converge to the objective world structure? Given our arguments so far, we suspect that this will happen under only a highly restricted set of circumstances. In the generic case, it seems unlikely that a sequence of Darwinian observers $\langle X_0, P_0 \rangle \rightarrow \langle X_1, P_1 \rangle \rightarrow \langle X_2, P_2 \rangle \rightarrow \ldots$ with monotonically increasing capacity of channel $P_j \colon W \rightarrow X_j$ for the expected fitness signal will automatically result in monotonically increasing channel capacity for the "truth" signal.

8.3 *Evolution of dedicated vs general-purpose representations*

We can conceive of the evolution of perceptual systems from "primitive" to "advanced" in a somewhat different way than above. In both options 8.1 and 8.2 considered above the specific fitness function $f_{o,s,a} \colon W \rightarrow \mathbb{R}$ was taken to be fixed. This meant that not only the organism o and its state s, but also the action class a, was assumed fixed. Both options above therefore considered the evolution of perceptual systems—either of the perceptual strategy P, or of the representation space X—for a specific fitness function associated with a fixed type of action (say, eating).

As organisms get more complex, however, their repertoire of action classes gets larger, and thus the number of specific fitness functions increases correspondingly. Evolution can deal with this increase in two possible ways. The first is that, as the number of action classes increases, the organism can evolve distinct representational spaces—one dedicated to each action class—in the sense of maximizing (or satisficing) the channel capacity for the

[7] Recall that a Darwinian observer is the combination of perceptual representational space Y, plus a perceptual channel P from the world W to Y.

expected fitness signal associated with that action class. There is certainly evidence that this happens in evolution; more advanced species tend to have a greater variety of perceptual representations, and many of these are highly dedicated to specific types of actions. It is likely, however, that at some point maintaining this strategy becomes infeasible; the number of distinct representations required may simply become too large.

The other possibility is for the organism to evolve representations that can subserve multiple types of actions at once. The perceptual strategy associated with such a—more general-purpose—representation will generally *not* maximize the channel capacity simultaneously for all of the expected fitness signals (one for each action class). However, it could achieve a high enough capacity for each expected fitness signal to make this general-purpose strategy worthwhile—especially since doing so obviates the need for multiple representational spaces.

Within this context, one can ask more specific questions:

(i) Under what conditions do selection pressures push in favor of one or the other solution: dedicated vs general purpose representations? Clearly the two are not mutually exclusive; indeed a single general-purpose representation that can subserve *all* action classes is extremely unlikely. A more realistic possibility is to have a small number of representational spaces, each of which subserves a cluster of action classes with somewhat similar fitness functions. The mathematics of how this would work in detail needs to be fleshed out (but see Shoval et al 2012 for a possible direction).

(ii) Consider a situation where a single representational space subserves an increasing number of action classes—hence an increasing number of fitness functions. Let X_k denote the evolving representational space that "accommodates" k fitness functions $f_1, f_2, \ldots, f_k$ (hence k different action classes, in the sense discussed above), and let $P_k\colon W \rightarrow X_k$ be the evolving perceptual channel corresponding to space X_k.[8] As the number of fitness functions k increases, it would appear that the evolving Darwinian observer $\langle X_k, P_k \rangle$ is tied less and less to any specific fitness function. Does increasing k therefore mean that P_k's capacity for the "truth" signal μ would also increase? In other words, are more "advanced" general-purpose representations more likely to capture objective world structure? Again, we suspect that this will not typically be the case; the general-purpose perceptual strategy will be a best-fit type solution relative to the given set of fitness functions $\{f_1, f_2, \ldots, f_k\}$ (eg similar to a least-squares regression line that passes as close as possible to a given set of data points). There is no principled reason why maximizing the channel capacity for the best-fit fitness signal should automatically maximize channel capacity for the truth signal. However, this must remain an open question until detailed mathematical models of this process are developed and studied.

9 General discussion

Marr's book *Vision* laid out an impressive and highly influential program of research in human vision—one that continues to motivate empirical research and theoretical models today. Indeed, the currently dominant BDT framework for vision may be viewed as a probabilistic approach to vision problems along essentially the lines laid out by Marr. BDT models vision at what Marr called the *computational level*—namely, an analysis of vision problems expressed in formal inferential terms, independently of how those inferences might be achieved by specific algorithms, or instantiated in neural hardware.

[8] In other words, the representational space X_2 accommodates two fitness functions $\{f_1, f_2\}$ corresponding to two different action classes, the representational space X_3 accommodates three fitness functions $\{f_1, f_2, f_3\}$ corresponding to three different action classes, etc.

In his treatment of vision, Marr seemed to make a principled distinction between human (or "advanced") vision and primitive vision. Specifically, he claimed that human vision generally computes objective properties of the world, whereas primitive visual systems compute only simple subjective properties—that nevertheless allow their owners to engage in specific, evolutionarily significant, behaviors. So whereas evolution shaped human vision to construct "a true description of what is there", it shaped primitive vision, such as that of the fly, so that it "does not really represent the visual world about it".

In the context of primitive vision, therefore, Marr granted that a visual system can fail to see any objective properties of the world and yet succeed at serving its owner adequately for the purposes of survival. In evolution, after all, what matters is fitness, not objective truth. But Marr also thought that as primitive visual systems evolve and become more complex, they move more and more towards computing objective properties of the world. Hence, with humans, he felt he could safely assume that the properties human vision computes are almost entirely objective.

BDT models of human vision share the assumption that human vision by and large sees the truth. Although they allow for misperceptions—in that the estimated value of a parameter might not match the "true" value—they assume, more fundamentally, that our language of perceptual representations is the right one for describing objective reality. Specifically, BDT assumes that the space of interpretations over which the posterior distribution is computed does contain somewhere within it a true description of the world. This assumption is closely tied to the historical roots of Bayesian approaches—namely, as a means of computing "inverse probability". In BDT applications to vision, this generally translates to treating vision as "inverse optics"—ie viewing the goal of vision as essentially "undoing" the effects of optics and "recovering" the scene that generated the projected image(s). It is in this more basic sense that BDT assumes that human vision sees the truth.

We have argued that in order to model human vision and its evolution, one must have a framework that is broad enough to allow for the possibility that we do not perceive the truth—not merely in the sense of missing the true estimate within the correct interpretation space, but in the more fundamental sense of not having the right language or representational framework for objective truth. Selection pressures in evolution are driven by fitness, not objective truth. Fitness is clearly distinct from truth because it depends not only on the objective world, but also on the *organism*, its *state*, and the *action class* in question. It is evident to us as humans that "lower" organisms such as flies and frogs do not have the capacity to represent objective truth—despite the fact that their evolved perceptual systems endow them with sufficient fitness to survive, even thrive. One's framework must allow for the possibility that the same is true of *Homo sapiens*—namely, that our evolved perceptual and cognitive representations do not provide us with the right language to understand objective truth (although they certainly provide an adequate representation for performing evolutionarily significant actions).

We outlined a framework, *Computational Evolutionary Perception* (CEP), that subsumes BDT and reinterprets it in evolutionary terms. CEP allows probabilistic inference much like BDT does; however, the perceptual inferences it draws are entirely between representational spaces (say, from a 2D representation of image structure to a surface representation that contains partial 3D information). The objective world W in CEP is unknown; we simply cannot assume that any of our perceptual inferences are being drawn in W. Whatever the nature of W, however, we do assume that: (i) it has some unknown probability distribution on it; and (ii) it has fitness functions defined on it. For a given organism o, state s, and action class a, specific fitness functions assign a fitness value to each element of W. Given a representational space X, evolution must tune the perceptual strategy or *channel* $P: W \to X$ to maximize—more correctly, satisfice—the capacity of the channel for the expected fitness signal. There is no mathematical reason to expect that increasing the channel capacity for the expected fitness signal will automatically result in increased capacity for the truth signal. The CEP framework also allows us to consider the evolution of representational spaces themselves, such that going from a simpler representational space (eg lower-dimensional or with less structure) to a more complex one (eg higher-dimensional or with more structure) confers substantially greater fitness. And CEP allows us to consider the evolution from primitive to advanced vision involving situations where, as the number of specific fitness functions increases (corresponding to an increase in relevant action classes), selection pressures may drive the evolution of more general-purpose representations that can subserve multiple types of actions. In each of these contexts, the CEP framework allows us to pose, and we hope eventually answer in formal terms, the question: Under what conditions, if any, do selection pressures drive perceptual systems to represent objective reality?

An overarching implication of the CEP framework is that the way we perceive the world reflects not objective reality, but the fitness functions that shaped natural selection (or, more accurately, it reflects a complex combination of objective reality and various fitness functions). Objects and surfaces, for example, feature prominently in the way that we perceive, and think about, the world. But there is nothing in modern physics that corresponds directly to things like tables and chairs. The fact that we evolved to represent the world in terms of objects and surfaces suggests, however, that this representational format probably made possible a high-capacity channel for expected fitness. Thus, according to the CEP framework: (i) objects are perceptual (and conceptual) representations, not entities that exist in objective reality (at least not in the way we intuitively think of them); and (ii) the reason why our perceptual representations of the world are organized in terms of objects is because such representations provided an effective coding scheme for expected fitness. *Thus, ultimately, objects are essentially an effective code for expected fitness.* The same applies to space itself: Objects are perceived not in isolation but as embedded in space, and with specific spatial relations between them. Our perceptual representations of space strongly emphasize such spatial relations. As with objects, it is likely that what we intuitively conceive of as space is another effective code for expected fitness.

We are not, of course, forever restricted to our perceptual representations. For example, our visual systems use stereo, motion, and other cues to estimate the depths of various objects. In this case, our perceptual scene interpretations are cast in a language involving 3D descriptions. However, our perceptions of 3D are clearly limited. When we look at the night sky, for instance, the stars look far away, but they all look roughly equally far away. From the depth perceptions of our visual system alone, we have no hint that two nearby stars could actually differ in depth from us by billions of miles.

The basic CEP framework in figure 3 showed only perceptual representations. In addition to these, we also have what we might call the "measured world", M—cognitive representations obtained as a result of applying scientific measurement procedures. When vision scientists say, for instance, that an observer underestimated the depth of an object, or misperceived its aspect ratio, they, of course, mean *with respect to spatial measurements taken in the physical (or a simulated) environment*. The measured world M thus clearly goes beyond our perceptual representations. At the same time, there is an important sense in which it is really an *extension* of our perceptual representations. For, although scientific procedures allow us to arrive at measurements that we could not possibly have arrived at from perception alone, the *formal structure* within which these measurements are placed is often dictated by the way we perceive the world, eg as having a three-dimensional structure. More precisely, we extend our 3D perceptions to a more adequate framework, M, often by making certain symmetry assumptions. For instance, if we assume that our 3D framework must be invariant under translations and rotations, then we might extend our perceptual representations of space to a Euclidean framework, as in Newtonian physics.

In this context it is natural to ask, "If our perceptions, and their extensions to the measured world, are not veridical, then how can we establish what the nature of the objective world really is?" The answer is to do normal science: Propose theories of the objective world, extract their empirical predictions, and test these predictions with experiments and simulations. One theory is that our perceptions of the objective world are normally veridical and that therefore the objective world resembles, in part, our perceptions. On evolutionary grounds, we doubt this theory. To reject it is not, however, to halt science. There are countless other theories to be explored and tested. Rejecting a false theory is genuine progress.

In sum, the proposed CEP framework captures the basic inferential structure of vision, subsumes BDT, and reinterprets it as evaluating expected fitness rather than estimating truth. It allows us to pose questions about the evolution of perceptual systems that would be difficult, if not impossible, to model within BDT. And it makes it relatively straightforward to incorporate higher-level conceptual representations, such as those based on scientific measurement, into the framework. The last point raises the intriguing possibility that many concepts in physics might be interpretable in terms of: (i) the evolution of perceptual representations based on maximizing—rather, satisficing—the channel capacity for the expected fitness signal; and (ii) using scientific measurements and symmetry assumptions to arrive at scientific concepts that extend our "raw" perception in various ways.

10 Mathematical appendix

For ease of reading we have kept mathematics to a minimum in the body of this paper. Here we fill in some details.

We assume that the *objective world* can be represented as a probability space $W(W, \mathcal{W}, \mu)$, where W is a set representing possible states of the objective world, $\mathcal{W}$ is a σ-algebra of events on W, and μ is a probability measure. We assume that *perceptions* of an organism can also be represented as a probability space $(X, \mathcal{X}, \mu_X)$, where X is a set of perceptions, $\mathcal{X}$ is a σ-algebra of events on X, and μ_X is a probability measure.

In the dispersion-free case, a *perceptual strategy* is a measurable function $P\colon W \rightarrow X$. This means that $\forall A \in \mathcal{X},\ P^{-1}(A) \in \mathcal{W}$, ie that the pullback of any perceptual event in $\mathcal{X}$ is a world event in $\mathcal{W}$. In this case, the probability measure μ on perceptions is related to the probability measure on the objective world as follows: $\forall A \in \mathcal{X},\ \mu_X(A) = \mu[P^{-1}(A)]$.

In the general case with dispersion, a *perceptual strategy* for a specific fitness function f is a Markovian kernel from W to X, ie a function $P\colon W{\times}\mathcal{X} \rightarrow [0, 1]$ such that (1) $\forall A \in \mathcal{X}$, $P(\cdot, A)$ is a measurable function on W; and (2) $\forall w \in W,\ P(w, \cdot)$ is a probability measure on X. In this case, the probability measure μ_X on perceptions is related to the probability measure μ on the objective world as follows:

$$\forall A \in \mathcal{X}, \quad \mu_X(A) = \mu P(A) = \int_W \mu(dw) P(w, A) .$$

The projection of the specific fitness function f onto X is the function f_X given by

$$f_X(x) = Pf(x) = \int_W P(w, dx) f(w) .$$

A perceptual strategy with dispersion, P, will not in general preserve structures on W. If, for instance, there is an order relation $\geqslant$ on W and, for some $w_1, w_2 \in W$, it happens that $w_1 \geqslant w_2$, then P will not in general map the indicator functions 1_{w_1} and 1_{w_2} to indicator functions of points in X, and thus will not preserve the order relation $\geqslant$. In this case one can develop a theory of the degree to which a structure is preserved, rather than insist that a structure be strictly preserved. In the special but interesting case in which P is an integral kernel of the form

$$P(w, A) = \int_A n[c(w), x] \lambda(dx) ,$$

where $A \in \mathcal{X}$, $c: W \to \mathbb{R}$ is a perceptual strategy without dispersion, $n: X \times X \to \mathbb{R}$ is a measurable function, and $\lambda: \mathcal{X} \to [0, 1]$ is a measure, then one can study which structures are strictly preserved by *c* and quantify in terms of *n* the degree to which those structures are preserved by *P*.

One structure that is always preserved by any perceptual strategy *P* is the Lebesgue order on the set M_w of all probability measures on (W, $\mathcal{W}$) (Bennett et al 1993).

Acknowledgments. For helpful discussions, we thank Jacob Feldman, Pete Foley, Geoff Iverson, Brian Marion, Justin Mark, Louis Narens, Darren Peshek, Negar Sammaknejad, and Kyle Stevens. We also thank Jacob Feldman for bringing the Einstein quote on the *object concept* to our notice. This research was supported in part by NIH grant EY021494 (MS) and a grant from Procter & Gamble (DH).

References

Adelson E H, Pentland A, 1996 “The perception of shading and reflectance”, in *Perception as Bayesian Inference* Eds D C Knill, W Richards (Cambridge: Cambridge University Press) pp 409–423

Bennett B M, Hoffman D D, Murthy P, 1993 “Lebesgue logic for probabilistic reasoning and some applications to perception” *Journal of Mathematical Psychology* **37** 63–103

Brainard D H, Freeman W T, 1997 “Bayesian color constancy” *Journal of the Optical Society of America A* **14** 1393–1411

Cover T M, Thomas J A, 2006 *Elements of Information Theory* (New York: Wiley)

Einstein A, 1936 “Physics and reality”, reprinted in *Ideas and Opinions* 1995 (New York: Modern Library) pp 318–357

Feldman J, 2012 “Tuning your priors to the world” *Topics in Cognitive Science* (in press)

Geisler W S, 2008 “Visual perception and the statistical properties of natural scenes” *Annual Review of Psychology* **59** 167–192

Geisler W S, Diehl R L, 2002 “Bayesian natural selection and the evolution of perceptual systems” *Philosophical Transactions of the Royal Society of London B* **357** 419–448

Geisler W S, Kersten D, 2002 “Illusions, perception and Bayes” *Nature Neuroscience* **5** 508–510

Glimcher P W, 2003 *Decisions, Uncertainty, and the Brain*: *The Science of Neuroeconomics* (Cambridge, MA: MIT Press)

Gregory R L, 1966 *Eye and Brain: The Psychology of Seeing* (London: Weidenfeld & Nicolson)

Gregory R L, 1970 *The Intelligent Eye* (London: Weidenfeld & Nicolson)

Gregory R L, 1974 *Concepts and Mechanisms of Perception* (London: Duckworth)

Helmholtz H L F von, 1910 *Treatise on Physiological Optics* translated by J Southal, 1925 (New York: Dover)

Hofbauer J, Sigmund K, 1998 *Evolutionary Games and Population Dynamics* (Cambridge: Cambridge University Press)

Hoffman D D, 2000 *Visual Intelligence: How We Create What We See* (New York: Norton)

Hoffman D D, 2009 “The interface theory of perception”, in *Object Categorization: Computer and Human Vision Perspectives* Eds S Dickinson, M Tarr, A Leonardis, B Schiele (Cambridge: Cambridge University Press) pp 148–165

Jameson K A, Komarova N L, 2009 “Evolutionary models of color categorization” *Journal of the Optical Society of America A* **26** 1414–1436

Jaynes E T, 2003 *Probability Theory*: *The Logic of Science* (Cambridge: Cambridge University Press)

Jepson A, Richards W, Knill D C, 1996 “Modal structure and reliable inference”, in *Perception as Bayesian Inference* Eds D C Knill, W Richards (Cambridge: Cambridge University Press) pp 63–92

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

Knill D C, Richards W (Eds), 1996 *Perception as Bayesian Inference* (Cambridge: Cambridge University Press)

Koenderink J J, 2011 “Vision and information”, in *Perception Beyond Inference. The Information Content of Visual Processes* Eds L Albertazzi, G V Tonder, D Vishnawath (Cambridge, MA: MIT Press) pp 27–58

Laplace P S, 1774 “Mémoire sur la probabilité des causes par les évènemens (Memoir on the probability of the causes of events)”, in *Mémoires de mathématique et de physique présentés à l’Académie royale des sciences, par divers savans, & lûs dans ses assemblées* **6** 621–656 (English translation by S M Stigler published in *Statistical Science* **1** 364–378, 1986)

Maloney L T, Zhang H, 2010 “Decision-theoretic models of visual perception and action” *Vision Research* **50** 2362–2374

Mamassian P, Landy M, Maloney L T, 2002 “Bayesian modeling of visual perception”, in *Probabilistic Models of the Brain: Perception and Neural Function* Eds R Rao, B Olshausen, M Lewicki (Cambridge, MA: MIT Press) pp 13–36

Mark J, Marion B, Hoffman D D, 2010 “Natural selection and veridical perceptions” *Journal of Theoretical Biology* **266** 504–515

Marr D, 1982 *Vision*: *A Computational Investigation into the Human Representation and Processing of Visual Information* (San Francisco, CA: Freeman)

Maynard Smith J, 1982 *Evolution and the Theory of Games* (Cambridge: Cambridge University Press)

Noë A, 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* Eds A Noë, E Thompson (Cambridge, MA: MIT Press) pp 567–598

Nowak M A, 2006 *Evolutionary Dynamics: Exploring the Equations of Life* (Cambridge, MA: Belknap Press)

Palmer S, 1999 *Vision Science: Photons to Phenomenology* (Cambridge, MA: MIT Press)

Pizlo Z, 2001 “Perception viewed as an inverse problem” *Vision Research* **41** 3145–3161

Prete F, 2004 *Complex Worlds from Simpler Nervous Systems* (Cambridge, MA: MIT Press)

Reichardt W, Poggio T, 1980 “Visual control of flight in flies”, in *Theoretical Approaches in Neurobiology* Eds W Reichardt, V Mountcastle, T Poggio (Cambridge, MA: MIT Press) pp 135–150

Sabra A I, 1978 “Sensation and inference in Alhazen’s theory of visual perception”, in *Studies in Perception: Interrelations in the History of Philosophy and Science* Eds P K Machamer, R G Turnbull (Columbus, OH: Ohio State University Press) pp 160–185

Samuelson L, 1997 *Evolutionary Games and Equilibrium Selection* (Cambridge, MA: MIT Press)

Sandholm W H, 2007 *Population Games and Evolutionary Dynamics* (Cambridge, MA: MIT Press)

Shepard R, 1994 “Perceptual–cognitive universals as reflections of the world” *Psychonomic Bulletin & Review* **1** 2–28

Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, Dekel E, Kavanagh K, Alon U, 2012 “Evolutionary trade-offs, Pareto optimality, and the geometry of phenotypic space” *Science* **336** 1157–1160

Singh M, Hoffmann D D, 2012 “Natural selection and shape perception: Shape as an effective code for fitness”, in *Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective* Eds S Dickinson, Z Pizlo (New York: Springer Verlag) (in press)