

# Evolutionary Dynamics and Accurate Perception. Critical Realism as an Empirically Testable Hypothesis

*Adriano Angelucci*

University of Urbino, Carlo Bo, Urbino (Italy)

*Vincenzo Fano*

University of Urbino, Carlo Bo, Urbino (Italy)

*Gabriele Ferretti*

Institute for Philosophy II, Ruhr-University Bochum (Germany); Eikones, University of Basel (Switzerland)

*Roberto Macrelli*

University of Urbino, Carlo Bo, Urbino (Italy)

*Gino Tarozzi*

University of Urbino, Carlo Bo, Urbino (Italy)

**Résumé :** Les modèles mathématiques peuvent être utilisés avec profit pour établir si notre perception du monde extérieur est précis. Donald Hoffman et ses collaborateurs ont développé un cadre mathématique prometteur dans lequel cette question peut être abordée, et qui repose sur une taxonomie exhaustive des différentes relations possibles qui peuvent tenir entre les représentations perceptuelles et le monde extérieur. Après avoir reformulé leur cadre au moyen d'un système formel amélioré, nous discutons de leur application de la théorie des jeux évolutifs, qui semble montrer qu'une stratégie perceptuelle essentiellement anti-réaliste à long terme surpasserait

---

biologiquement ses rivaux. Nous soutenons que leur modèle ne prend pas dûment en considération la signification biologique cruciale des changements environnementaux et proposons des modèles alternatifs qui le font. Nous concluons que, dans nos modèles, une représentation partiellement réaliste serait privilégiée.

**Abstract:** Mathematical models can be profitably used to establish whether our perception of the external world is accurate. Donald Hoffman and his collaborators have developed a promising mathematical framework within which this question can be addressed and which is based on an exhaustive taxonomy of the different possible relations between perceptual representations and the external world. After reformulating their framework by means of an improved formal system, we discuss their application of evolutionary game theory, which appears to show that an essentially anti-realistic perceptual strategy would in the long run biologically outcompete its rivals. We argue that their model does not take the crucial biological significance of environmental changes into due consideration and propose alternative models which do. We conclude that a partially realistic representation would be favoured in our models.

## 1 Introduction

The following is arguably a fundamental question about human cognition: Does natural selection tend to favour individuals whose perceptual systems are epistemically reliable? Over the last few decades, different mathematical models have been used to address this question. It should be noted, however, that our question raises two different, albeit related issues, the first psychobiological and the second epistemological. The former is descriptive in nature and concerns the kind of perceptual strategies available to complex organisms under the pressure of natural selection. The latter is normative and concerns the cognitive evaluation of those same strategies. We will accordingly begin by addressing the epistemological issue raised by Donald Hoffman and his collaborators (henceforth, H & C, see, in particular [Hoffman, Singh *et al.* 2015], [Mark, Marion *et al.* 2010], but also [Hoffman, Singh *et al.* 2013] and [Hoffman & Singh 2012]) and then move on to engage closely with the current debate on perceptual strategies in evolutionary theory.

We will briefly consider Godfrey-Smith's approach [Godfrey-Smith 1991] in order to show that it *presupposes* perceptual accuracy rather than investigating it. We will then discuss a new proposal from H & C, which posits the idea that our perception is tuned to fitness and is *therefore* completely unreliable. Their approach to this issue will be the main topic of our paper. We intend to show that, contrary to their view, the platitude opining that fitness is one

of the most important forces responsible for shaping our perceptual strategy does not imply that our perception is altogether unreliable. What led them astray, we suggest, is the fact that their model fails to take the relevance of environmental modifications into due account. A model that does take these into account, as we will try to show, integrates the idea that the acquisition of apparently useless information about the environment in fact increases fitness, at least up to a certain point. Moreover, according to the same model, even when two organisms make use of the same number of bits, modifications of the environment will in general favour the one that uses its information to target reality rather than utility. As a consequence, perceptual strategies aimed at acquiring reliable information about the environment will in the long run be favoured over those aimed at increasing utility.

The plan is as follows. Section 2 briefly presents Godfrey-Smith and H & C's approaches, section 3 sketches a theoretical framework within which our fundamental question can be stated in a philosophically exact manner. Section 4 tackles what is arguably the most powerful argument levelled by H & C against critical realism, namely one based on evolutionary game theory. In section 5 we put forward two alternative models, which assume that fitness covaries with the environment surrounding the cognitive agent. Then, we use our model to show that, in a biologically more realistic situation, knowledge of the objective structure of the world proves to be a rather valuable asset. We conclude by evaluating our results in section 6 and by listing the technical notions and equations we rely on in Appendices A and B.

## 2 The accuracy of perception

In 1991 Peter Godfrey-Smith developed an interesting framework based on signal theory to model the optimal relation between the inner states of an organism and the external world. His approach assumes the reliability of what he calls the “world-head” relation—i.e., the organism's dispositions to act in response to a given stimulus—and it statically represents the relation between that stimulus and the inner state. The same framework was later generalized and couched in terms of Bayesian decision theory [see, e.g., Knill & Richards 1996]—i.e., by relying on Bayesian conditionalization to model the way in which the mind evaluates hypotheses about the distal stimulus based on its perception of the proximal one. This approach builds on a distribution of an a priori probability of the possible realities of a distal stimulus and an a priori distribution of likelihood of a certain perception with respect to those realities. Then, through Bayes' conditionalization rule, the a posteriori probability distribution of the hypotheses corresponding to those diverse realities can be established. Within this broader framework, it is possible to study the role played by evolution in perception, as the framework allows for the idea that in the long run organisms will modify

their a priori probability distributions in accordance with a certain utility function determined by biological fitness. One could hence try to establish whether this function will produce a priori probability distributions in the long run that would favour perceptual accuracy. Yet, to the best of our knowledge, this line of research has never been systematically pursued and philosophical reflections on the Bayesian approach [cf. e.g., Rescorla 2015] seem to presuppose the accuracy of perception rather than investigating it. For this reason, we will set these proposals aside in this paper and leave scrutiny thereof for a later occasion.

In a series of articles (see, in particular, [Hoffman, Singh *et al.* 2015], [Mark, Marion *et al.* 2010], but also [Hoffman, Singh *et al.* 2013] and [Hoffman & Singh 2012]), Donald Hoffman and his collaborators (henceforth, H & C) have constructed a controversial, yet apparently powerful mathematical case, which gives a negative answer to our question. The currently mainstream view among scientists studying perception is that, in normal conditions, our perceptual representations are largely accurate.<sup>1</sup> Had this not been the case—it is usually argued—our species would have long been driven to extinction [Geisler & Diehl 2003], [Yuille & Bülthoff 1996]. This is also true for every kind of biological system. In order to question this widely held assumption, H & C appeal to evolutionary games, which purportedly show that our visual<sup>2</sup> systems are tuned to utility rather than objective reality and that we therefore have little or no reason to believe that our perceptual representations are ever, or even broadly accurate. The gist of their evolutionary argument is that individuals whose visual systems integrated this last property would badly lose an evolutionary competition to individuals whose visual representations were tuned to utility rather than objective reality. Their proposal has the undeniable merit of lending empirical meaning to various philosophical views concerning the adequacy and accuracy of our perceptual systems. The proposal is hence deeply thought-provoking both from scientific and philosophical standpoints.

In this paper we will take issue with some of the arguments defended by H & C and put forward a new model, whose epistemological implications run counter to those purportedly supported by theirs. Our model suggests that—contrary to H & C’s view—an individual whose visual representations were at least partially accurate would be more successful from an evolutionary point of view. In line with contemporary philosophical theories of perception [Nanay 2013], [Siegel 2006], [Brogaard 2014], [Ferretti & Zipoli Caiani 2019] we assume

---

1. For instance, [Marr 1982, 340], [Trivers 2011, 2], and [Pizlo, Li *et al.* 2014, 227]. H & C’s views are couched in terms of *veridicality* rather than *accuracy*. We prefer the latter term, as we assume that the similarity relation holding between a content and its reference is a matter of degree rather than a yes-or-no affair.

2. H & C focus on the case of visual perception, as they assume that the same will hold for other sense modalities. We will hence follow suit and focus on visual perception only.

that perception is a representational process,<sup>3</sup> i.e., that it represents reality by ascribing various features to objects as well as to the visual scene as a whole.<sup>4</sup> Our main goal is to provide a plausible and empirically testable model of visual perception, standardly conceived of as a process whereby a stimulus causally interacts with our visual system in turn giving rise to a more or less accurate representation of its source, i.e., a *perception*. It is generally acknowledged that perceptual representations have accuracy rather than truth conditions [Mausfeld 2002], [Siegel 2006, 2010] and [Koenderink 2014]. As a consequence, this paper will discuss the accuracy of our visual representations. It should be noted that the question we are facing is not “whether there is biologically useful information in the environment” as this is assumed to be the case. The important question is in fact “whether it is biologically useful to acquire this information, even if its acquisition is biologically costly”.

### 3 Epistemological definitions based on the “best possible theory”

Following H & C partially, let us consider an environment characterized by a set  $\mathbf{R}$  of features, every subset of which can be a stimulus and can therefore cause a corresponding subset of the set  $\mathbf{P}$  of possible contented mental states in a biological sensitive being. Then let us consider the best possible theory  $\mathbf{T}$  of these two aspects of reality. The notion of a “best possible theory” is necessary to face our problem, since we are investigating the accuracy of the representation of the environment by a perceptive biological being. To evaluate this accuracy, we need an effective description of the two *relata* we are comparing, namely the environment and its representation. Moreover, the supposition of a best theory  $\mathbf{T}_{\mathbf{R}}$  of the stimulus is not a *petitio principii*. Indeed, it is possible to object that we are investigating the accuracy of perception and assume that we know how the world is constituted but if perception is not at entirely accurate it is very improbable that we could possess this kind of knowledge. Nonetheless the epistemological part of our

---

3. Despite various attempts at defending a nonrepresentational view of perception [Noë 2004], [Chemero 2009], [Hutto & Myin 2013], representationalism is still the dominant view of perception, mainly due to its undeniable explanatory advantages [Pautz 2010], [Nanay 2013]. It is clear that, if perception were direct even in a weak sense, H & C would be a fortiori wrong.

4. We are aware that there are many different philosophical theories of perception. However, as already noticed by the neurophysiologist Johannes Müller in the nineteenth century (cf. [Boring 1950, chap. 5]) and as abundantly confirmed by subsequent empirical research (cf. [Kandel, Schwartz *et al.* 2013, 415–416]), the phenomenal qualities of our perceptual representations depend on the neurophysiological pathways responsible for elaborating perceptual stimuli. As a consequence, we will assume that representationalism is the best theory of perception available at the moment.

paper—and the same was true for H&C—is not devoted to establishing a proactive tool to measure the accuracy of perception. We instead prefer to outline a general theoretical framework which works to define exactly the problem we are investigating on. For this reason, presupposing a best possible theory of the stimulus is an idealization which is allowed.

**T** will include two sub-theories—namely **T<sub>R</sub>** and **T<sub>P</sub>**—which deal with **R** and **P** respectively.<sup>5</sup> If these two theories are developed enough, we will also have two corresponding state-spaces **S<sub>TR</sub>** and **S<sub>TP</sub>**. Roughly speaking, if a theory is based on  $n$  features, its state-space is  $n$ -dimensional and one of the features is represented on each dimension. The notion of “state-space of a theory” is useful, since the set of possible states of a system—according to the theory—is a sub-region of the state-space. Moreover, it is easy to build a correspondence between the two state-spaces of **T<sub>R</sub>** and **T<sub>P</sub>** and then evaluate its accuracy.

We will begin by briefly sketching the way in which H&C set out the problem of evaluating the accuracy of perception. We will then modify it in order to render it epistemologically more appropriate.<sup>6</sup> However our modifications will not affect H&C’s main point which will be discussed in the subsequent sections.

Hence, leaving details aside, we can imagine a representation function **F** mapping **S<sub>TP</sub>** regions onto **S<sub>TR</sub>** ones—i.e., roughly, a model of our capacity to represent a stimulus given a perception—and an inverse causal function **Q** mapping **S<sub>TR</sub>** regions onto **S<sub>TP</sub>** ones—i.e., roughly, a formal representation of the causal link between stimuli and contented mental states. A *perceptual strategy*<sup>7</sup> can now be defined as the composite function **FQ** mapping **S<sub>TR</sub>** regions onto **S<sub>TR</sub>** ones. A given strategy may be dubbed *realist* if **FQr<sub>SR</sub>** = **r<sub>SR</sub>** (where **r<sub>SR</sub>** is a region of **S<sub>TR</sub>**)—i.e., if our perceptions perfectly mirror the stimuli that cause the corresponding contented states. As it does not seem plausible to suggest that biological perceptive systems generally implement this kind of strategy, it is useful to define a further strategy that we may call *critical realist*. A strategy is critical realist if there is at least a subspace of **S<sub>TR</sub>** within which realism holds—i.e., within which, if **r<sub>S'R</sub>** is a region of **S'<sub>TR</sub>** and **S'<sub>TR</sub>**  $\subset$  **S<sub>TR</sub>**, then **FQr<sub>S'R</sub>** = **r<sub>S'R</sub>**. If, on the other hand, **S'<sub>TR</sub>** =  $\emptyset$ , then the resulting strategy will coincide with what H&C call *interface theory*.

H&C distinguish an initial causal process which goes from the stimulus to the experience of the biological system and a further process which goes from the representation of the biological system to reality. In this way

5. We assume, for the sake of simplicity, that it is possible to distinguish between the two theories, although we are aware that they are connected.

6. Thanks to an anonymous referee, who emphasized this problem.

7. Note that here the term “perceptual strategy” introduced by H&C does not involve any forms of behaviour of the organism. “Strategy” refers only to the accuracy relation between environment and its representation.

the evaluation of accuracy of perception becomes particularly simple, since the stimulus and its representation are in the same state-space. However, this model is not appropriate, because the causal and the representational processes are the same phenomenon seen from two different points of view. The former is seen from the perspective of the mind and the latter from the perspective of the environment. Indeed, there are *not* two different mental contents—what the biological system perceives and what the biological system considers real—but only one, namely the representation of the object caused by the impingement of the latter in the sense organs of the system. This means that accuracy must be evaluated comparing the  $\mathbf{T}_P$  and the  $\mathbf{T}_R$  state-spaces, that is  $\mathbf{S}_{TP}$  and  $\mathbf{S}_{TR}$ , not comparing  $\mathbf{S}_{TR}$  before perception with  $\mathbf{S}_{TR}$  after perception, as proposed by H & C, which is a much more difficult task. In order to overcome this problem, we developed a mathematical framework, the details of which can be found in Appendix A. Here we shall briefly illustrate its result. According to the model, it is possible to measure the accuracy of perception of a biological system through the *distance* between the objective conjunct probability measure on the state-space of the stimuli  $\mu_{RR}$  and the conjunct probability measure induced by the perception on the state-space of sensations  $\mu_{PP}$ .

In this context, it will not be necessary to go into the details of how to measure the distance between these two functions. It is indeed possible to define an inner product between probability measures and then an Euclidean distance. This distance ( $d_{RP}$ ) can be normalized so that when two measures are the same its value is “0” and when two given elements belonging to an algebra defined on  $\mathbf{S}_{TR}$   $x_R$  and  $y_R$  are such that “ $\mu_{PP}(x_P, y_P) = 1 - \mu_{RR}(x_R, y_R)$ ”<sup>8</sup> its value is 1.

At this point it is easy to establish that a *critical realist* strategy determines a value of  $d_{RP} \leq 0.5$ , whereas an *interface strategy* determine a value of  $d_{RP} > 0.5$ .

Going a bit deeper into the question of measuring the distance between  $\mu_{RR}$  and  $\mu_{PP}$ , it is reasonable to suppose that there are sub-regions of  $\mathbf{S}_{TR}$  such that  $d_{RP}$  is almost 1 and others where it is  $< 0.5$ . We may also suppose there to be cases in which the perception is quite accurate and others in which it is quite inaccurate. If this is true, one can define a “critical realist” perceptual strategy as involving a *representative* set of cases in which  $d_{RP}$  is near 1. It is “representative” in the sense that for each different kind of sensitive nervous terminations there is at least a perceptive *normal* situation in which  $d_{RP}$  is almost 1. Obviously, if this does not hold true then perception follows an “interface strategy”.

Let us now consider the main argument that H & C rely on in order to show that *Homo sapiens* implements an interface strategy (henceforth, **IF**).

---

8. See Appendix A for further explanations. In any case, this equation means that representation and reality are altogether different.

## 4 Evolutionary game theory

Evolutionary game theory is arguably the best way to predict the evolution of a discrete phenotypic trait whose fitness depends on its frequency in the population [Rice 2004, 263]. Since we are discussing the evolution of different perceptual strategies which can coexist, the fitness of a strategy will depend on the frequency of the other.

Mark, Marion *et al.* imagine a simple situation, in which an organism finds itself in an environment divided in three territories [Mark, Marion *et al.* 2010, henceforth, M&C]. Each territory has one resource, whose quantity takes discrete values from 1 to 100. Utility (which corresponds to fitness) is proportional to the quantity of resource present on the territory. They calculate the payoff for a realist strategy, i.e., a strategy that gathers all the available information, as well as of a critical realist strategy, i.e., a strategy based on only two possible perceptions, according to whether the quantity of the resource is above, equal or below a certain threshold  $\beta$ . In calculating the payoffs, the authors assume that the interface strategy will always provide a choice first, as the realist strategy will take longer to gather information (and the time spent will have a negative weight which is proportional to the quantity of acquired bits). Their calculations apparently show that the payoff for the realist strategy (assuming a cost of 1 per bit) will drive realism to extinction for any value of  $\beta$ . The same would happen if we increased the complexity of an environment. This result was indeed to be expected, given that the realist strategy was never very plausible to begin with as we observed above.

M & C next consider a game in which 3 strategies compete, namely realism, critical realism and **IF**. At this point, though, instead of relying on the approach adopted in the preceding case—which could indeed be applied to a  $3 \times 3$  matrix—the authors write:

[...] unlike in two-player games, we cannot immediately calculate which strategy survives and which becomes extinct. Instead we compute the time derivatives of the frequencies of each strategy using the replicator equation. [Mark, Marion *et al.* 2010, 513]<sup>9</sup>

The replicator equation is known to be of huge importance and influence for the study of the evolutionary system in biology [Cressman & Tao 2014] and [Nowak 2006].

The article reports no calculations as such and just a triangular simplex,<sup>10</sup> whose vertexes represent populations implementing only one of the three available perceptual strategies. Inner points correspond to mixed populations

9. This equation can be used to calculate the time derivative of the frequency of a certain trait, based on its fitness, as well as on the average fitness of all the traits, see [Rice 2004, 282].

10. Every point in the simplex represents a possible value of the three frequencies, whose value must add up to 1.



whose overall dynamics (represented by means of arrows) *point toward the interface strategy from most points of the simplex*, in case information has reasonable energetic costs.

It is interesting to analyze the playing field of the three strategies in this case. As in the previous strategy, we have 3 territories and one resource whose values range from 1 to 100. Yet now the utility is a Gaussian with its peak at 50 and is not proportional to the quantity of the resource to be found on each territory anymore. As in the previous case, the realist strategy gathers all the information and the critical realist strategy relies on 3 perceptions (e.g., 3 colours). The same is true for the interface strategy. The difference between the two latter strategies lies in how colours are used (Figure 1).

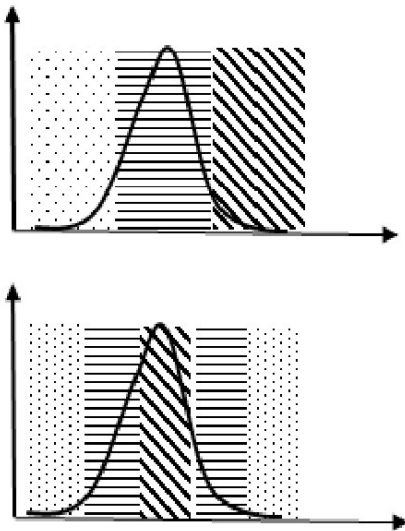


Figure 1 – Critical realist strategy (above); Interface strategy (below). On the  $x$ -axis the quantity of resource; on the  $y$ -axis the utility. The difference between critical realist and interface strategies is expressed through diverse distributions of colours. The different colours are represented through the diverse types of filling: points, horizontal lines and diagonal lines. It is evident that the use of colours in the interface strategy is more useful—in terms of fitness—than its counterpart in the critical realist strategy.

The quantity and utility of a resource are represented on the  $x$  and  $y$ -axes respectively. While the critical realist strategy simply associates the three colours with the increasing quantity of the resource, disregarding the utility curve, the interface strategy only keeps track of utilities. As a consequence, the latter graph is divided in three non-adjacent zones, which track utilities rather than quantity. When quantity and utility are non-monotonically related, each

strategy must be charged with the costs associated with gathering information and calculating its corresponding utility. The authors modulate the cost of knowledge to make it comparable to the cost associated to knowledge of utility.

This stage has been clearly and intentionally set to put critical realism at a disadvantage, since, by keeping perceptual complexity fixed,<sup>11</sup> the interface strategy has a running start. For instance, a situation where **IF** is disadvantaged was pointed out by Anderson [Anderson 2015]: organisms often tend to homeostasis and hence they have an interest in knowing whether the quantity of a given resource is above or below a certain threshold. In this kind of case, the critical realist strategy would have an advantage over its interface counterpart. However, H & C’s model is biased for a more profound reason.

To demonstrate the limits of this example, let us now consider the following case. Suppose that the organisms in question are sparrows and the resource little worms. Worms evolve, so we can imagine a random mutation that significantly decreases their size. Since sparrows tend to catch the bigger worms, the new trait spreads rapidly across the population. This considerably alters the ratio between the utility of the resource and its quantity. The sparrow implementing an interface strategy will accordingly still “think” that the same number of worms is needed in order to maximize utility and hence will lag behind in terms of fitness. Its critical realist competitor will “know better” and will accordingly move to an area where either more worms or bigger worms are to be found. We take this case to show that slight modifications in the environment can bring about a serious disadvantage for the interface strategy. Indeed if a perceptual strategy completely tuned with utility in a certain environmental scenario does not also gather some apparently useless information, it is in danger of not “understanding” the importance of possible environment modifications in terms of utility.

## 5 Our models

The present section is divided into two parts. In its first part, we consider a case in which two different perceptual strategies—critical realist and interface—have the same number of bits available. This is a situation similar to the one considered in the preceding section. In its second part, we consider a case in which a critical realist strategy collects more information than an interface one.

In the first situation, the difference between the two strategies is that they use bits in different ways—i.e., whereas the critical realist strategy attempts

---

11. The authors propose situations in which the stage is different, but they never consider the case in which the environment changes.

to at least partially represent the environment, the interface strategy uses information in order to target utility only.<sup>12</sup>

Our main point is that, given a static environment, a strategy targeting utility will clearly outcompete one aimed at representing reality. In a case where the environment changes, the opposite will be true however. We will now try to show this by means of a very simple model according to which an interface strategy will *initially* prevail over a critical realist one and yet this trend will reverse in due time and because of modifications in the environment.

Let us then consider an organism of genus  $\omega$ , itself divided into two species  $\omega_{CR}$  and  $\omega_{IF}$  which differ in perceptual strategy only. Consider next two environmental features— $x$  and  $y$ —interacting in a way such that, as time goes by, their global utility for  $\omega$  fluctuates. The quantities of  $x$  and  $y$  will fluctuate as well—i.e., the phases of their oscillations will be such that their overall utility will fluctuate as well.

Let us imagine a situation similar to that of beats, where the sum of two sinusoids which are partially out of phase (measured by  $\varepsilon$ ) will form a sinusoid with a smaller frequency. The quantity of  $x$  and  $y$  could, e.g., be given by:

$$x = a \cos t, \quad y = b \cos \varepsilon t.$$

We assume that, as time passes without genetic mutations but with the natural cycle of the environment, the objective utility function for  $\omega$  will be:

$$U = \left( a \cos t + b \cos \varepsilon t - \frac{ab}{I} \cos t \cos \varepsilon t \right).$$

This means that  $U$  will indeed oscillate as time goes by. However, it will not be the mere sum of  $x$  and  $y$ , but rather a smaller quantity which decreases inversely with respect to constant  $I$ . While our utility function appears quite specific, it is in fact more general than may seem the case at first sight. In this regard, it should be noted that that  $U$  covaries with  $x$  and  $y$ . It indeed seems reasonable to assume that there will be a function connecting utility to environmental features.  $U$ 's third term, in particular, establishes that the connection between  $x$  and  $y$  is non-linear. This term becomes more relevant as  $I$  increases—i.e., when  $I$  is high, the difference at time 0 between interface and critical strategy becomes smaller. This means that, in a static environment, an interface strategy will prove more useful. All in all, it seems quite natural to suppose that environmental features will interact in a non-linear way. Imagine, for instance, that an organism's calorie intake comes either from lipids or from carbs. It seems reasonable to expect that its utility will not be proportional to the mere sum of these two resources, as it will also presumably depend on their relative quantity. Moreover, it is also

---

12. In a first version of the paper we neglected this very important case. We would like to thank an anonymous referee for bringing this point to our attention.

reasonable to expect that an organism tracking utility has higher fitness in a static environment. On the contrary, in a changing environment, a critical strategy will prove more useful.

Note also that for  $t = 0$ , the following will hold:

$$U = a + b - \frac{ab}{I}.$$

We define  $U_{IF}$  ( $U_{CR}$ ) as the received utility function by  $\omega_{IF}$  ( $\omega_{CR}$ ), given its peculiar way of encoding information. Let us now suppose that  $U_{IF}$  does not change with time and is given by the following formula:

$$U_{IF} = a + b - \frac{ab}{I}.$$

On the contrary,  $\omega_{CR}$  knows the value of  $x$  and  $y$  with the same accuracy as  $\omega_{IF}$  evaluates  $a$  and  $b$ . This means that  $\omega_{CR}$  and  $\omega_{IF}$  will have the same quantity of bits available.

Since  $\omega_{CR}$  knows  $x$  and  $y$ , it can evaluate a utility function of this kind:

$$U_{CR} = x + y = a \cos t + b \cos \varepsilon t.$$

That is the mere sum of the two variables.

At this point the distances between the real utility and the utility evaluated by the two strategies can be defined in the following straightforward way:

$$d_{IF} = |U - U_{IF}| \text{ and } d_{CR} = |U - U_{CR}|.$$

Note that at  $t = 0$   $U_{CR} = a + b$ . This means that at  $t = 0$ ,  $d_{CR} > d_{IF} = 0$   $t = 0$ , that is  $U_{IF}$  is better than  $U_{CR}$ . The crucial questions thus become: What will happen as time goes by? Which strategy will perform better? It is easy to calculate that  $d_{IF} > d_{CR}$  for  $t > 0$ .

According to our model, then, when the environment is held fixed and each organism has the same number of bits at its disposal, a strategy aimed at increasing utility will outcompete one aimed at representing reality. As the environment changes, however, the opposite will be the case.

While our model is admittedly limited in scope, the assumptions upon which it is based seem quite reasonable. We take those assumptions are the following:

1. Environment changes.
2. Many environmental features display an oscillating pattern.
3. Utility is not in general the mere sum of two such features.
4. A constant utility function is not appropriate to represent utility in a changing environment.

5. Knowledge of the environmental features' behaviour, while itself insufficient to locate the real utility function, nonetheless seems a reasonable starting point to assess utility in a changing environment.

In light of the above, we maintain that generally in a changing environment a critical realist perceptual strategy will outperform an interface strategy, even in a case in which two organisms have the same number of bits available.

Let us now move on to the second situation—the one in which the critical realist strategy collects more bits than the interface one.

Most mathematical literature on evolution focuses on the way in which a phenotypic or genotypic trait evolves given a certain fitness [Rice 2004]. This can change on some models over time due to modifications in the environment [Gillespie 1991, chap. 4], whereas our goal is to investigate the impact of environmental changes<sup>13</sup> on fitness variations of a biological system. To our knowledge, no mathematical studies in this direction have been carried out to date. Our very simple model focuses on the variation of fitness due to changes in the environment, without modifying genotype and phenotype of the population. To simplify our calculation, we presuppose asexual reproduction and no mutations. Moreover, the environment can change discretely generation by generation.

Let us consider a set  $\{\omega_K\}_{K=0,\dots,N}$  of populations differing only in their capacity to acquire knowledge from the environment. Imagine that a population  $\omega_K$  of a certain species at generation  $i$  has absolute fitness<sup>14</sup>  $w_i$ . This population lives in a complex environment that can be modelled by means of  $N$  possible, biologically relevant dichotomic changes concerning a certain set of given environmental features. Obviously, the number of features an organism can collect is smaller than  $N$  and the relation between  $N$  and this number is given by the discrete capacity of environmental discrimination of  $\omega_K$ . No mutation occurs in the sequence of generations. Suppose that  $\omega_K$  has a quantity  $K$  ( $0 \leq K < N$ ) of information about the features—equivalent to  $K$  bits—and that acquiring  $K$  information about each one of them has an increasing cost  $Kc/(N - K)$  in terms of fitness, where  $c$  is an average unit of cost. We use this kind of cost because when  $\omega_K$  acquires little information the cost is low but when the information obtained grows the cost increases sharply due to the excessive time employed.<sup>15</sup> In this case, the cost associated with

---

13. The relation between an organism and its environment is an extremely complex subject matter. On this point we shall follow [Lewontin 2000], according to whom physical traits become biologically relevant only with respect to a given organism while organisms constantly modify their environment. It should also be noted that approach modifications in the environment play a central role in evolution on the punctuated equilibria.

14. Number of offspring.

15. We are aware that this model seems to reveal nothing about how knowledge of fitness-relevant features (i.e.,  $f$ ) of the world is acquired. However introducing this kind of cost would underline the fact that acquisitions of more and more information increases the cost per unit of information.

the acquisition of information gets bigger and bigger. When  $K$  goes to  $N$ , the harm suffered by  $\omega_K$  tends to 0 but increases probabilistically as  $K$  gets smaller. On the other hand, if  $K$  increases, the cost for gathering information becomes bigger and bigger; if  $K = N$ , the cost would be infinite.

After a generation, the environment will have undergone a certain number of changes,<sup>16</sup> only an amount  $f$  of which are biologically relevant. Let us assume—plausibly enough—that most of these changes are either irrelevant or harmful, hence, in particular, most changes that have an impact on fitness will be slightly harmful.<sup>17</sup> The probability distribution that  $\omega_K$  lacks information about each and every change will be uniform. Since “ $N - K$ ” is the number of bits unknown by  $\omega_K$  and the whole number of relevant information is  $N$ , then the uniform distribution gives  $(N - K)/N = 1 - K/N$  for each information. Moreover we assume that the newborns have to gather all knowledge useful for their fitness obtained and also that there is no costless transmission of knowledge from one generation to the following. Let’s say that the average negative impact of these changes on fitness is  $w_-$ . Therefore, the fitness  $w_{i+1}$  of generation  $i + 1$ , given the fitness  $w_i$  of generation  $i$ , is given by the following difference equation (1):

$$w_{i+1} = w_i - f(1 - K/N)w_- - Kc/(N - K). \quad (1)$$

$Kc/(N - K)$ , that is the third term of the right part of (1), is the cost for population  $\omega_K$  of gathering  $K$  bits of information about its environment, whereas the second term is the biological damage for  $\omega_K$  due to its ignorance. This equation should tell us which value of  $K$  in  $\omega_K$  will prove advantageous over generations, in a changing environment. By solving the equation, (the details of which can be found in Appendix B), we define  $K_{max}$  as

$$K_{max} \equiv N \left( 1 - \sqrt{\frac{c}{fw_-}} \right). \quad (2)$$

$K_{max}$  (2) establishes the best value of  $K$  in order to postpone extinction.<sup>18</sup> To understand whether our model is a good argument favouring a critical realist perceptual strategy, we must compare (2) with our epistemological definition in the preceding section. The distance  $d_{RP}$  between the conjunct probability measures on  $\mathbf{S}_{TR}$  and  $\mathbf{S}_{TP}$  depends on the number of faithful causal connections imposed by  $\mathbf{Q}$ . We assume that the faithfulness of perception of one feature is independent from that of another. This is

16. Each change could be represented by 1 bit.

17. Here we follow [Eldredge 1998], according to whom important changes in the environment are a fundamental engine of evolution, as they favour the extinction of most species present in a given place.

18. An anonymous referee criticized our choice only to consider extinction. Yet, as is well known, 99% of the species appeared on the Earth in the history of life became extinct.

reasonable, since there are no clear reasons for connecting the faithfulness of knowledge of different features. Let us assume that the probability distribution of ignorance on information is uniform. Then the probability of not knowing a piece of information about a feature is given by  $(1 - K/N)$ . Therefore the rate of success of conjunct probability measures in **STP** is proportional to  $K^2/N^2$ . In this case  $d_{RP}$  is equal to zero because the two measures,  $\mu_{PP}$  and  $\mu_{RR}$ , are the same; otherwise  $d_{RP}$  is proportional to  $1 - K^2/N^2$ . Now, when returning to (2), we note that if  $f$ , the number of harmful environmental changes, is high, then  $K_{max}$  tends to  $N$ , that is  $K^2/N^2$  tends to 1 and the distance  $d_{RP}$  tends to 0, that is a critical realistic perceptual strategy is favoured from an evolutionary point of view. To sum up, among the  $\{\omega_K\}_{K=0,\dots,N}$  populations we investigated, the  $\omega_{\sim N}$  (the population with  $K \approx N$ ) is favoured. And the  $\omega_{\sim N}$  population has a critical realist perceptual strategy.

Clearly this result is only partial. In a biologically more realistic situation, one would also have to consider the non-deleterious aspects of the environment. Moreover, objective laws of the environment are not constituted only by association of features, but rather as more complex relations. Finally, different organisms could deploy the usefulness of the acquired information in various degrees. A much more complete study would be needed to obtain more conclusive proof that accuracy is a good goal to pursue from a selective point of view.

## 6 Concluding remarks

If perception is the only way to acquire information about our environment and perception turns out not to be even partially accurate, then investigating *Homo sapiens* and its environment would amount to inspecting our subjectivity. Yet modern science's moral and cognitive mission also consists of pursuing fallible and revisable attempts at formulating justified hypotheses about *Homo sapiens*, its origins and the world it inhabits. Many cultural milieus encourage the idea that empirical science cannot make any justified claims about the external world, independently of the way in which that world is perceived or thought of. If perception were completely inaccurate, this idea would be reinforced. Whether and the extent to which human perception is an accurate representation of the world is an epistemological matter which can be empirically investigated at least indirectly by using evolutionary mathematical models. We showed the limits of H & C's attempts at establishing the negative impact on fitness of an accurate representation of the world. We also used a simple model to show that, in a more plausible biological situation, the at least partial accuracy of an organism's perceptual representations proves evolutionarily successful.

Our model is clearly only in sketch form at this stage and certainly requires further development. Indeed, we are confident that, given reasonable assumptions concerning what should count as an accurate perception, it should

be possible to empirically investigate the comparative fitness of different perceptual strategies along the lines suggested by H&C. However we also believe that such investigation should carefully take into account modifications in the environment.

## Appendix A

We can suppose that  $\mathbf{T}_R$  is a physical theory and  $\mathbf{T}_P$  a theory from the psychology of perception. In establishing this comparison between  $\mathbf{S}_{TR}$  and  $\mathbf{S}_{TP}$  we will have two problems: The *first* is that the features of  $\mathbf{T}_R$ —wave-lengths, masses, objective distances, etc.—are completely different from the features of  $\mathbf{T}_P$ —colours, weights sensations, perceived distances, etc.; the *second* is that we must find a criterion to connect causally a point of  $\mathbf{S}_{TP}$  to a point of  $\mathbf{S}_{TR}$ . In other terms, how can we know that a certain change in colour seen by a biological system is the effect of a certain change in the world?

Before confronting these difficulties, we establish that in our setting, a *perceptual strategy* is simply the causal relation between  $\mathbf{S}_{TR}$  and  $\mathbf{S}_{TP}$  which is the mapping  $\mathbf{Q}$  from sub-regions of  $\mathbf{S}_{TR}$  onto sub-region of  $\mathbf{S}_{TP}$ .  $\mathbf{R}$  is the part of reality which is perceptually relevant for a given biological system. Therefore the dominion of  $\mathbf{Q}$  is all  $\mathbf{S}_{TR}$ . Clearly  $\mathbf{Q}$  is not one-to-one, since two different stimuli can be perceived as equal. Moreover  $\mathbf{Q}$  is not onto either, since the biological system can perceive something that is not real (hallucination).

As is the case in H&C's framework, from our perspective it is easy to define what is direct or naïve realism.  $\mathbf{Q}$  is a *direct realistic* strategy if it is a mapping from  $\mathbf{T}_{SR}$  to  $\mathbf{T}_{SP}$  only for a very small sub-region of  $\mathbf{S}_{TR}$  and for the most part of  $\mathbf{S}_{TR}$  it is an identity in  $\mathbf{S}_{TR}$ . Like H&C, we assume that this is not a biologically reasonable perceptual strategy.

To investigate more realistic perceptual strategies, we assume that on  $\mathbf{S}_{TR}$  it is possible to establish a  $\sigma$ -algebra  $\mathbf{A}_R$ , that  $\mathbf{S}_{TR}$  has a good structure to be a sample space of a probability measure. Let us call the latter  $\mu_R$ . Now, take two generic events belonging to  $\mathbf{A}_R$  and let us call them  $x_R$  and  $y_R$ .  $\mu_R(x_R)$  and  $\mu_R(y_R)$  are both probability measures, therefore also:

$$\mu_{RR} = \mu_R(x_R \wedge y_R) = \mu_R(x_R)\mu_R(y_R/x_R) = \mu_R(y_R)\mu_R(x_R/y_R)$$

is a probability measure.

Let us assume that  $\mathbf{Q}$  applied to  $\mathbf{A}_R$  imposes a  $\sigma$ -algebra  $\mathbf{A}_P$  on  $\mathbf{S}_{TP}$ . Now, let us consider the probability distribution  $\mu_P$  on this imposed algebra. If  $x_R$  and  $y_R$  belong to  $\mathbf{A}_R$ , let us define  $x_P = Q(x_R)$  and  $y_P = Q(y_R)$  respectively. If  $\mu_R(x_P)$  and  $\mu_R(y_P)$  are probability measures, then also:

$$\mu_{PP} = \mu_P(x_P \wedge y_P) = \mu_P(x_P)\mu_P(y_P/x_P) = \mu_P(y_P)\mu_P(x_P/y_P)$$



is a probability measure.

Now, a brief explanation is in order. The probability distribution  $\mu_{RR}$  is a good representation of how couples of real features behave insofar as  $\mu_{RR}$  at least describes partially what happens objectively. It is clear that to have a more complete objective description of stimuli we should consider more complex aggregates of features and not just couples of these. However in this context, to consider only couples is enough to clarify the epistemological problem we are working with.

“Q” answers to our second problem, i.e., Q establishes the causal connections between events in  $S_{TR}$  and events in  $S_{TP}$ .

On the contrary, to solve our first problem, namely the inhomogeneity between  $S_{TR}$  features and  $S_{TP}$  features, we compare the two conjunct probabilities distributions  $\mu_{PP}$  on  $S_{TP}$  and  $\mu_{RR}$  on  $S_{TR}$  rather than  $x_P$  with  $x_R$  and  $y_P$  with  $y_R$ . It should be noted that  $\mu_{PP}$  is induced on  $S_{TP}$  by Q applied to events in  $S_{TR}$ .

## Appendix B

To solve the equation (1), we get (3):

$$w_i = w_0 + im \tag{3}$$

where  $m = -fw_-(1 - K/N) - Kc/(N - K)$  and  $w_0$  is the initial fitness of  $\omega_K$ .

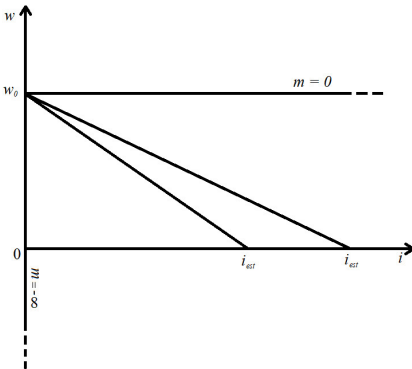


Figure 2 – On the  $x$ -axes the number of generations; on the  $y$  the fitness. Representation of the family of intersecting straight lines of equation (1). More  $i_{est}$  are drawn to show how their value can vary when  $K$  increase or decrease.

Representing the equation’s solutions as  $m$ ’s slope changes (Figure 2), we get a family of intersecting straight lines that pass through the point  $P(0, w_0)$ .

Step  $i$  corresponds to  $\omega_K$ 's extinction (call it  $i_{est}$ ). This result, which is obtained when  $w_i = 0$ , has the general form:

$$i_{est} = \frac{w_0}{fw_- \left(1 - \frac{K}{N}\right) + \frac{Kc}{N-K}}. \tag{4}$$

The bigger  $m$  is, the greater the extinction step. The extinction values (4) covary with  $K$ . Remember that we are considering a set of populations with different rates of knowledge represented by the number  $K$ . Notice that the value of  $m$  will certainly be non-positive insofar as the non-negative values of  $f, w_-, c, N$  and  $K$  vary. This happens to be why we are concerned only with harmful environmental modifications for the sake of simplicity. Moreover,  $m$  cannot obtain the full range of negative values, but only those obtained by letting  $K$  vary from 0 to  $N$ , i.e., we consider only those populations having from 0 knowledge to all relevant knowledge. Reordering the expression for  $m$ , we get (5):

$$m = \frac{Kfw_-}{N} - \frac{Kc}{N-K} - fw_-.$$

We can now see that  $m$  increases as  $K$  gets bigger if:

$$K < N \left(1 - \sqrt{\frac{c}{fw_-}}\right),$$

assuming  $c \leq fw_-$  to have a value of  $K \geq 0$ .

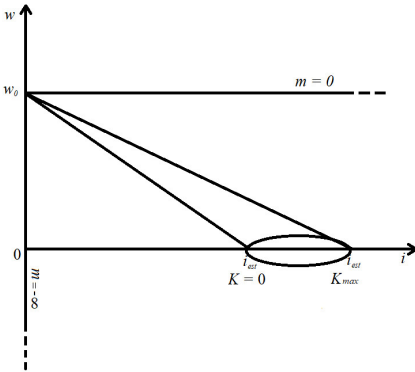


Figure 3 – The values of  $i_{est}$  which interest us are those between  $K = 0$  and  $K$  corresponds to the maximum value of  $m(K_{max})$  (the circled segment).

If we set<sup>19</sup>  $c = w_-$ , we see that  $m$ 's values get bigger for  $K$  if:

$$K < N \left( 1 - \sqrt{\frac{1}{f}} \right).$$

it should be noted that  $1 \leq f \leq N$ , i.e., we assume that there is at least 1 deleterious variation of the environment. Hence, if  $f$  increases,  $K_{max}$  (2) tends to omniscience ( $K \rightarrow N$ ), i.e., the values of  $K$  for which  $m$  increases are all except  $N$ . Recall that as  $m$  increases,  $i_{est}$  gets bigger. On the other side, if  $f$  is small,  $K \rightarrow 0$ , i.e., it is not relevant for the fitness of  $\omega_K$  to gather information. This means that, if the environment changes significantly, it pays to have information. For instance, if at least 1 out of 10 changes in the environment is harmful for a given organism, information, even if partial, will be convenient. Indeed, in this case we have:

$$K < N \left( 1 - \sqrt{\frac{10}{N}} \right).$$

One can see that if  $N$ —i.e., the number of biologically relevant information of the environment—is big—as is plausible—the values of  $K$  for which  $i_{est}$  increases are very high with respect to  $N$ .

To sum up, the number of harmful environmental changes will be very high for fairly complex organisms and it will thus be reasonable to suppose that  $\omega_K$  with values of  $K \approx N$  is favoured, i.e., populations with a good knowledge of their environment. What our simple model shows, then, is that a population with a high  $K$  will win out in a changing environment.

## Acknowledgements

We would like to warmly thank Colin Allen and two anonymous reviewers for addressing important comments, both on the experimental and on the philosophical side, which led us to significantly improve the first version of this article. Gabriele Ferretti has some specific acknowledgments to do. Since the finalization of the work for the publication of the present article has been done during the transition from one research position to another, Gabriele Ferretti would like to acknowledge support from both the corresponding fellowships. This work was supported by a NOMIS Fellowship, awarded by the Eikones-Center for the Theory and History of the Image at the University of Basel, Switzerland. This work was also supported by a Humboldt Fellowship, hosted by Professor Albert Newen at the Institute for Philosophy II, Ruhr-University Bochum, Germany.

---

19. Which is the value which allows interface to win in M & C

## Bibliography

- ANDERSON, Barton L. [2015], Where does fitness fit in theories of perception?, *Psychonomic Bulletin & Review*, 22(6), 1507–1511.
- BORING, Edwin Garrigues [1950], *A History of Experimental Psychology*, New York: Appleton, 2nd edn.
- BROGAARD, Berit (ed.) [2014], *Does Perception Have Content?*, New York: Oxford University Press.
- CHEMERO, Anthony [2009], *Radical Embodied Cognitive Science*, Cambridge; London: MIT Press.
- CRESSMAN, Ross & TAO, Yi [2014], The replicator equation and other game dynamics, *Proceedings of the National Academy of Sciences*, 111(Supplement 3), 10 810–10 817, doi: 10.1073/pnas.1400823111.
- ELDREDGE, Niles [1998], *Pattern of Evolution*, New York: Freeman.
- FERRETTI, Gabriele & ZIPOLI CAIANI, Silvano [2019], Between vision and action: introduction to the special issue, *Synthese*, Special issue, online, doi: 10.1007/s11229-019-02518-w.
- GEISLER, Wilson S. & DIEHL, Randy L. [2003], A bayesian approach to the evolution of perceptual and cognitive systems, *Cognitive Science*, 27(3), 379–402, doi: 10.1207/s15516709cog2703\_3.
- GILLESPIE, John H. [1991], *The Causes of Molecular Evolution*, New York: Oxford University Press.
- GODFREY-SMITH, Peter [1991], Signal, decision, action, *The Journal of Philosophy*, 88, 709–722, doi: 10.2307/2027008.
- HOFFMAN, Donald D. & SINGH, Manish [2012], Computational evolutionary perception, *Perception*, 41(9), 1073–1091, doi: 10.1068/p7275.
- HOFFMAN, Donald D., SINGH, Manish, *et al.* [2013], Does evolution favor true perceptions?, in: *Human Vision and Electronic Imaging XVIII*, edited by B. E. Rogowitz, T. N. Pappas, & H. de Ridder, International Society for Optics and Photonics, SPIE, vol. 8651, 10–21, doi: 10.1117/12.2011609.
- [2015], The interface theory of perception, *Psychonomic Bulletin & Review*, 22(6), 1480–1506, doi: 10.3758/s13423-015-0890-8.
- HUTTO, Daniel D. & MYIN, Erik [2013], *Radicalizing Enactivism: Basic Minds without Content*, Cambridge, Mass.; London: MIT Press.

- KANDEL, Eric R., SCHWARTZ, James H., *et al.* [2013], *Principles of Neural Sciences*, New York: McGraw Hill.
- KNILL, David C. & RICHARDS, Whitman [1996], *Perception as Bayesian Inference*, Cambridge: Cambridge University Press.
- KOENDERINK, Jan [2014], The all seeing eye?, *Perception*, 43(1), 1–6, doi: 10.1068/p4301ed.
- LEWONTIN, Richard [2000], *The Triple Helix*, Cambridge, Mass.: Harvard University Press.
- MARK, Justin T., MARION, Brian B., *et al.* [2010], Natural selection and veridical perceptions, *Journal of Theoretical Biology*, 266(4), 504–515, doi: 10.1016/j.jtbi.2010.07.020.
- MARR, David [1982], *Vision*, San Francisco: Freeman.
- MAUSFELD, Rainer [2002], The physicalistic trap in perception theory, in: *Perception and the Physical World*, edited by D. Heyer & R. Mausfeld, New York: Wiley, 75–112.
- NANAY, Bence [2013], *Between Perception and Action*, Oxford: Oxford University Press.
- NOË, Alva [2004], *Action in Perception*, Cambridge, Mass.: MIT Press.
- NOWAK, Martin A. [2006], *Evolutionary Dynamics. Exploring the Equations of Life*, Cambridge, Mass.: The Belknap Press.
- PAUTZ, Adam [2010], Why explain visual experience in terms of content, in: *Perceiving the World*, edited by B. Nanay, New York: Oxford University Press, 254–309, doi: 10.1093/acprof:oso/9780195386196.003.0010.
- PIZLO, Zygmunt, LI, Yunfeng, *et al.* [2014], *Making a Machine that Sees Like Us*, New York: Oxford University Press.
- RESCORLA, Michael [2015], Bayesian perceptual psychology, in: *The Oxford Handbook of Philosophy of Perception*, edited by M. Matthen, Oxford: Oxford University Press, 694–716, doi: 10.1093/oxfordhb/9780199600472.013.010.
- RICE, Sean H. [2004], *Evolutionary Theory. Mathematical and conceptual foundations*, Sunderland: Sinauer.
- SIEGEL, Susanna [2006], Which properties are represented in perception?, in: *Perceptual Experience*, edited by T. S. Gendler & J. Hawthorne, Oxford: Oxford University Press, 481–503, doi: 10.1093/acprof:oso/9780199289769.003.0015.

— [2010], *The Contents of Visual Experience*, New York: Oxford University Press.

TRIVERS, Robert [2011], *The Folly of Fools*, New York: Basic Books.

YUILLE, Allan L. & BÜLTHOFF, Heinrich H. [1996], Bayesian decision theory and psychophysics, in: *Perception as Bayesian Inference*, edited by D. C. Knill & W. Richards, Cambridge University Press, 123–162, doi: 10.1017/CBO9780511984037.006.