

Structural coding versus free-energy predictive coding

Peter A. van der Helm¹

Published online: 25 September 2015
© Psychonomic Society, Inc. 2015

Abstract Focusing on visual perceptual organization, this article contrasts the free-energy (FE) version of predictive coding (a recent Bayesian approach) to structural coding (a long-standing representational approach). Both use free-energy minimization as metaphor for processing in the brain, but their formal elaborations of this metaphor are fundamentally different. FE predictive coding formalizes it by minimization of prediction errors, whereas structural coding formalizes it by minimization of the descriptive complexity of predictions. Here, both sides are evaluated. A conclusion regarding competence is that FE predictive coding uses a powerful modeling technique, but that structural coding has more explanatory power. A conclusion regarding performance is that FE predictive coding—though more detailed in its account of neurophysiological data—provides a less compelling cognitive architecture than that of structural coding, which, for instance, supplies formal support for the computationally powerful role it attributes to neuronal synchronization.

Keywords Cognitive architecture · Free-energy minimization · Neuronal synchronization · Perceptual organization · Predictive coding · Structural coding

Introduction

The term “predictive coding” is nowadays often used to refer to a family of models of perceptual inference in the hierarchically organized visual cortex. It is a diverse family, with various Bayesian and artificial neural network models, some of which can process images while others cannot. It is also a family with a divide regarding the roles of feedforward and feedback connections. One set of models assume that feedback connections carry predictions while feedforward connections carry prediction errors (e.g., Rao & Ballard, 1999). The other set of models assume that feedforward connections carry predictions while feedback connections carry constraints on these predictions (e.g., Lee & Mumford, 2003). In this article, I contrast a recent Bayesian model from the former set of models to a long-standing representational model close to the latter set of models.

More specifically, focusing on theoretical aspects, I contrast Friston’s (2009, 2010) Bayesian version of predictive coding to the representational approach called structural coding (see van der Helm, 2014). Like other predictive coding models, these two models aim at unifying competence (i.e., what is a system’s output?) and performance (i.e., how does the system arrive at this output?). Special is that both models use free-energy minimization as metaphor for processing in the brain, but with totally different elaborations of this metaphor. One difference is that, in free-energy (FE) predictive coding, predictions are based on probabilities, whereas in structural coding, they are based on descriptive complexities.¹ This is basically merely a difference

✉ Peter A. van der Helm
peter.vanderhelm@ppw.kuleuven.be
perswww.kuleuven.be/peter_van_der_helm

¹ Laboratory of Experimental Psychology,
University of Leuven (K.U. Leuven), Tiensestraat 102,
Box 3711, Leuven 3000, Belgium

¹ Descriptive complexities and probabilities represent formal quantifications of the strength or quality of things, that is, they are not taken to be represented in the brain as actual numbers.

in means, but it led, in these two coding approaches, to fundamentally different views on hierarchical perceptual inference. The core ideas of the two coding approaches may be introduced briefly as follows.

FE predictive coding, on the one hand, draws on von Helmholtz's (1909/1962) idea, also known as the likelihood principle, that we perceive the most likely objects or events that would fit the sensory input that we are trying to interpret (cf. Hochberg, 1978; Gregory, 1973; Pomerantz & Kubovy, 1986). Strong versions take "most likely" to refer to objective probabilities in the world (which does not seem tenable; Feldman, 2013; van der Helm, 2000, 2011), but Bayesians usually take it to refer to subjective probabilities, or beliefs. In any case, FE predictive coding assumes that the visual system tests predictions in a top-down fashion—along recurrent (or feedback, or reentrant, or descending) neural connections—against the sensory input. Prediction errors are returned in a bottom-up fashion—along feedforward (or ascending) connections—to update the to-be-recycled predictions (Bastos et al., 2012). This process is driven by prediction-error reduction, which is seen as reflecting free-energy minimization (Friston, 2009, 2010) and which is formulated in terms of Shannon's (1948) classical information theory.

Structural coding, on the other hand, draws on the Gestalt law of Prägnanz (Koffka, 1935; Köhler, 1920, 1929; Wertheimer, 1912, 1923). This law was inspired by the idea that the brain, like any physical system, tends to settle in relatively stable states defined by a minimum of free energy. It is generally understood to refer to a tendency towards regularity, symmetry, and simplicity, or as Koffka (1935) formulated it for vision: "*Of several geometrically possible organizations that one will actually occur which possesses the best, the most stable shape*" (p. 138). Building on this Gestaltist idea and on seminal work by, for instance, MacKay (1950), Hochberg and McAlister (1953), Attneave (1954), and Garner (1962), structural coding began as a competence model (Leeuwenberg, 1968), but nowadays, it also includes performance (van der Helm, 2012, 2014, 2015a).

Structural coding assumes that the perceptual process in the visual hierarchy in the brain comprises three neurally intertwined subprocesses, namely, feedforward extraction of visual features from sensory input, horizontal (or lateral) binding of similar features, and recurrent selection of different features to be integrated into percepts (cf. Lamme, Supèr, & Spekreijse, 1998). These three subprocesses, together, are assumed to yield simplest hierarchical organizations of sensory input—that is, organizations in terms of wholes and parts, which are formally definable by a minimum number of descriptive parameters. This idea—which reflects Occam's razor—is also known as the

simplicity principle. It is in line with modern information theory, which may need some further introduction.

Modern information theory arose in reaction to Shannon's (1948) classical information theory. Whereas classical information theory requires knowledge of actual probabilities to optimize things, modern information theory aims to do more or less the same without needing to know actual probabilities. Since the 1960s, it developed into algorithmic information theory (AIT) in mathematics (see Li & Vitányi, 1997), and independently, into structural information theory (SIT) in human visual perception research (see Leeuwenberg & van der Helm, 2013). There are differences between AIT and SIT, but currently relevant, they share the Occamian idea that the simplest interpretation of data is the best one² (for discussions on these issues, see van der Helm, 2000, 2011, 2014).

Hence, the idea that visual perception is a form of unconscious inference governed by free-energy minimization is a long-standing Gestaltist idea adopted first by structural coding and only later by FE predictive coding (which, to my knowledge, has been silent about these historical roots). However, the two coding approaches differ fundamentally regarding the questions of (a) how this idea is cast in information-theoretic terms, and (b) what the underlying neural mechanisms entail. The former question is a competence question, which I address first—also because it provides a leg up to the latter question, which is a performance question. To focus on the broader conceptual issues, I skip technical details of both coding approaches (these can be found elsewhere). For the same reason, I elaborate neither on the wealth of neurophysiological data that is claimed to support FE predictive coding (see Clark, 2013), nor on the wealth of behavioral data that is claimed to support structural coding (see Leeuwenberg & van der Helm, 2013).

Competence

As said, whereas predictions in FE predictive coding are based on probabilities, predictions in structural coding are based on descriptive complexities. However, a descriptive complexity C can be converted into the artificial probability $p_a = 2^{-C}$, which is called an algorithmic probability in AIT (Li & Vitányi, 1997) and a precisal in SIT (van der Helm, 2000). It reflects that simpler things are

²Solomonoff (1964a, 1964b) and Kolmogorov (1965) initiated AIT by proving that simplicity is a fairly stable concept, that is, it does not matter much which descriptive coding language is used. Independently, Simon (1972) observed this in perception, when he compared six representational coding models.

assigned higher probabilities, and implies that structural coding can be given a Bayesian formulation too (see Fig. 1). Before discussing this further, it is expedient to contrast this modern information-theoretic notion of precisal to the classical information-theoretic notion of surprisal (term by Tribus, 1961), which plays a role in FE predictive coding (where it is called “surprise”).

The precisal, on the one hand, is a probability derived from a descriptive complexity, that is, from an information quantification based on a description of an individual message (e.g., a hypothesis), whose hierarchical internal structure reflects that of the message. The surprisal, on the other hand, is Shannon’s (1948) solution to get an optimal encoding of messages, that is, to minimize the long-term average burden on communication channels given the transmission probabilities of pre-chosen messages. The surprisal of a message is the negative logarithm of its transmission probability relative to those of all other possible messages, and optimal encoding is achieved by labeling all messages with arbitrary nominalistic codes the length of their surprisals. Thus, more likely messages are assigned shorter labels (as, e.g., in the Morse Code). There is some debate in mathematics whether precisals form a proper probability distribution, but notice that the surprisal is definitely not a descriptive complexity: It is an information quantification based on a message’s probability and is unrelated to the message’s internal structure. Hence, as van der Helm (2000, 2011) argued earlier, it is factually incorrect to claim that the two information quantifications are formally equivalent,

as either implicitly or explicitly has been claimed by Chater (1996), Friston (2010), and Thornton (2014), for instance.

Bayesian modeling

Bayes’ rule (Bayes & Price, 1763) is a powerful mathematical modeling tool given by:

$$p(H|D) = \frac{p(H) * p(D|H)}{p(D)}$$

In words, Bayes’ rule holds that, for data D to be explained, the posterior probability $p(H|D)$ of hypothesis H is proportional to the prior probability $p(H)$ of H , multiplied by the conditional probability $p(D|H)$ of D if H were true. The probability $p(D)$ of D is the normalization factor. In general, Bayesian approaches aim to establish a posterior probability distribution over the hypotheses, but a specific goal is to select the most likely hypothesis, that is, the one with the highest posterior probability under the employed prior and conditional probabilities. To formulate this specific goal, the normalization factor $p(D)$ can be omitted, yielding:

Select the H that maximizes $p(H|D) = p(H) * p(D|H)$

In perceptual organization, Bayes’ rule can be applied to determine the posterior probability $p(H|D)$ of a candidate interpretation H of sensory data D . Such an interpretation, or scene model, comprises a hypothesized organization of the distal stimulus, that is, it comprises hypothesized distal objects that could fit the sensory data. The prior $p(H)$ then is the probability of interpretation H independently of sensory data D , that is, it can be said to indicate how good hypothesis H is in itself (it is therefore also said to account for view-independent properties of H). Furthermore, the conditional $p(D|H)$ then is the probability of sensory data D if interpretation H were true, that is, it can be said to indicate how well data D fit hypothesis H (it is therefore also said to account for view-dependent properties of H).

In FE predictive coding, hypotheses are assumed to be given beforehand, and prediction errors are defined by conditional surprisals, that is, by the negative logarithm of $p(D|H)$. So, in classical information-theoretic terms, it aims to minimize the surprisal of data D given hypothesis H . In structural coding, conversely, hypotheses are assumed to be constructed on the fly from the sensory data, and in modern information-theoretic terms,³ it aims to minimize the sum of (a) the prior complexity of hypothesis H , and (b) the conditional complexity of data D given hypothesis H (Fig. 2 gives a gist). In Bayesian terms, FE predictive

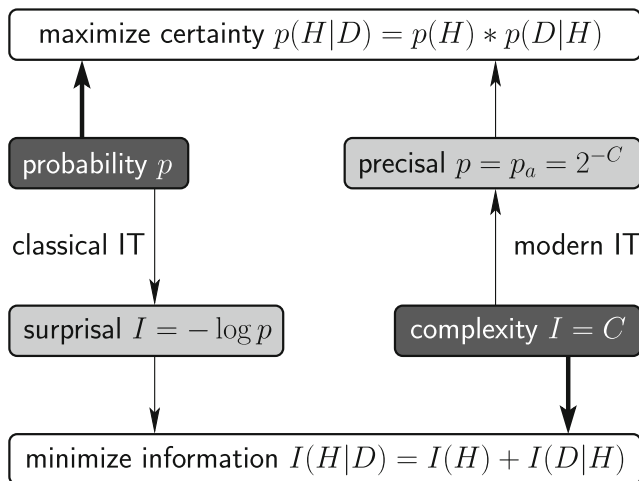


Fig. 1 Objective or subjective probabilities p can be used to maximize Bayesian certainty, and via the surprisal conversion from classical information theory (classical IT), also to minimize information as quantified in classical IT. Descriptive complexities C can be used to minimize information as quantified in modern information theory (modern IT), and via the precisal conversion from modern IT, also to maximize Bayesian certainty under these probabilities

³As formulated by Rissanen (1978) in AIT, and independently, by van Lier et al. (1994) in SIT.

coding aims to maximize conditional probabilities,⁴ while structural coding aims to maximize the product of prior and conditional precisals.

Furthermore, as said, in Bayesian models, probabilities usually are beliefs, that is, probabilities based on an individual's past experience, or knowledge,⁵ while in structural coding, they are precisals, that is, probabilities derived from descriptive complexities. Notice that precisals can be said to reflect a belief but that not every belief is reflected by precisals. This may seem obvious, but as I discuss next, it yet deserves clarification.

No automatic inclusion of Occam's razor

Bayes' rule is a selection method, whereas Occam's razor, or the simplicity principle, is a selection criterion—just as the Helmholtzian likelihood principle. Bayesian models can accommodate any selection criterion, so, also Occam's razor. However, there is a persistent misconception that every Bayesian model agrees automatically with Occam's razor. This misconception seems to have arisen in the early 1990s, when it also got its first refutation by Wolpert (1995). It reappeared in Chater's (1996) claim, reiterated by Feldman (2009), that the simplicity and likelihood principles are equivalent, which was refuted by van der Helm (2000, 2011). Nevertheless, invoking Chater (1996) and Feldman (2009), Thornton (2014) persisted—crucially flawed in that he ignored the fundamentally different ways in which classical and modern information theory quantify information (see the beginning of section “Competence”). Furthermore, just as Feldman (2009), Thornton (2014) invoked an argument by MacKay (2003), which van der Helm (2011) had refuted as follows.

MacKay argued that a category of more complex instances spreads probability mass over more instances than a category of simpler instances does, so that such simpler instances tend to get higher probabilities. Notice that this presupposes (a) a correlation between complexity and category size, and (b) that every category gets an equal probability mass. These presuppositions are inherent neither to Bayes' rule nor to the Helmholtzian likelihood paradigm. In fact, they are at the heart of the following, insightful, reasoning about the reliability of simplicity as predictor.

⁴FE predictive coding is silent about what the priors (i.e., the probabilities of hypotheses H independently of data D) might be, even though these too are needed for a full Bayesian account. Friston (2010) included optimization of what he called empirical priors, but these empirical priors are not independent of the sensory data (see Section “Empirical priors”; see also Trappenberg & Hollensen, 2013).

⁵Neurons responding to only certain stimuli, or features relevant according to structural coding, do not fall under this definition of knowledge. They rather fall under the sensitivity of a system to certain input—just as a sieve is sensitive to (but not knowledgeable about) things of only certain sizes.

Imagine a world with objects generated by, each time, first selecting randomly a complexity category, and then selecting randomly an instance from that category. Thus, in the first step, all categories have the same probability of being selected, and in the second step, all instances in the selected category have the same probability of being selected. By definition, instances in a category of complexity C are describable by C parameters, so, the category size is proportional to 2^C . This implies that the probability that a particular instance is selected is proportional to 2^{-C} —which, notably, is the earlier-mentioned precisal p_a . Hence, in this particular kind of world—which MacKay seemed to have in mind—the simplicity and likelihood principles are equivalent, but notice that this says nothing about how these principles are related in other imagined or actual worlds.

In other words, MacKay's argument is not an argument that Bayesians can use to claim automatic inclusion of Occam's razor, but it is one that Occamians might use to promote Occam's razor as a belief worthy of building Bayesian models on. In AIT, this belief has been supported by showing, among other things, that simplest descriptive codes yield near-optimal encoding if the actual probability distribution is one from the infinite set of enumerable probability distributions (Li & Vitányi, 1997). Thus, simplest descriptive codes can be said to have a general-purpose nature in that they yield fairly optimal encoding in many imaginable worlds.⁶ As I discuss in a moment, something similar holds for the veridicality of simplest descriptive codes.

Hence, Bayesian models can comply with Occam's razor, but they do not comply automatically with it. To comply with Occam's razor, one would have to start from precisals or, if one prefers to use objective probabilities, one would have to assume a world like the one MacKay (2003) apparently had in mind. As far as I can tell, this holds for all types of Bayesian models. That is, it holds for both parametric Bayesian models (in which predictions depend on chosen belief parameters, as, e.g., in FE predictive coding) and non-parametric Bayesian models (where “nonparametric” means that belief parameters are adjusted on the fly as the incoming data are gathered; for such models of cognition, see, e.g., Austerweil & Griffiths, 2013). It also holds for hierarchical Bayesian inference models, which I discuss in the section on performance, because they are intimately related to ideas about neural implementation. By way of prelude to this, but

⁶Recall that optimal encoding minimizes the long-term average burden on information channels. Shannon achieved this by an efficient labeling of all messages with nominalistic codes, which are unrelated to the internal structure of individual messages. Here, near-optimal encoding is achieved by minimal coding of the internal structure of individual messages, which seems a more appropriate reflection of the way in which the brain might encode sensory information in an efficient and parsimonious fashion.

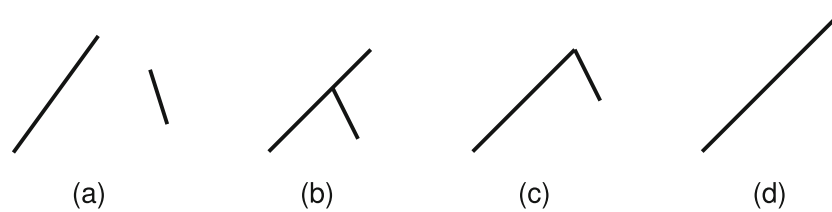


Fig. 2 Each of these configurations can be interpreted as consisting of a long segment and a short segment. The prior complexity of this two-objects hypothesis reflects the effort to construct these segments, and the conditional complexities reflect the effort to bring the segments in each of the given positions. For details of the quantification of conditional complexities, see van Lier et al. (1994), but it corresponds roughly to the intuitively assessed number of positional degrees

of freedom to be removed to arrive at a given position. This implies it increases gradually in (a)–(d). In (c) and (d), the relatively high conditional complexities imply that the one-object hypothesis is predicted to prevail, while in (b), the two-objects hypothesis is predicted to prevail (confirmed by Feldman, 2007). The latter agrees with common ideas that such T-junctions between the contours of two shapes are cues that one shape occludes the other

still pertaining to competence, I next discuss plain Bayesian inference.

Bayesian inference and the role of action in perception

Bayesian inference is basically the recursive application of Bayes' rule. In perceptual organization, this general technique is particularly convenient to model visual updating by moving observers, as van der Helm (2000) explicated as follows (see also Fig. 3).

A moving observer usually gets a growing sample D of different views (i.e., proximal stimuli) of the same distal scene. Suppose sample D consist, at first, of only one view, with H_i ($i = 1, 2, \dots$) as candidate interpretations and with prior and conditional probabilities $p(H_i)$ and $p(D|H_i)$, so that the posterior probabilities $p(H_i|D)$ can be determined by applying Bayes' rule. Then, each time an additional view enters the sample D , the previously computed posterior probabilities $p(H_i|D)$ can be taken as the new prior probabilities $p(H_i)$ which, together with the conditional probabilities $p(D|H_i)$ for the expanded sample D , can be used to determine new posterior probabilities by again applying Bayes' rule. This recursive application of Bayes' rule is not guaranteed to converge always on one interpretation (cf. Diaconis & Freedman, 1986), but generally, it converges on one interpretation, which, under the employed conditionals, will continue to get the highest posterior when sample D is expanded further (cf. Li & Vitányi, 1997).

Hence, if one has (approximately) the right conditional probabilities, then several (not too atypical) views of a distal scene suffice to make a (fairly) reliable inference about what the distal scene comprises and, thereby, what subsequent views will show. That is, the trick of the recursive application of Bayes' rule is that, after several recursions, the effect of the first priors fades away because the priors are updated continuously on the basis of the conditionals, which, thereby, become the decisive entities. This useful trick brings me to the next two observations on the role of action in perception.

First, AIT found that the margin between precisals and probabilities from an enumerable probability distribution P is maximally equal to the complexity of P (Li & Vitányi, 1997)—this complexity corresponds roughly to the number of categories to which P assigns probabilities. This holds for priors and conditionals, and again illustrates the general-purpose nature of simplest descriptive codes, which, by this finding, can be said to be fairly veridical in many imaginable worlds. For perception, this can be sharpened as follows. The number of prior categories in the world is very high, so that prior precisals are probably not very veridical. However, the number of conditional categories for a specific hypothesis is relatively small, that is, there usually are few qualitatively different views of a scene—this suggests that conditional precisals are pretty veridical. For instance, if one throws two sticks on the floor, then the result might be one of the four configurations in Fig. 2—with, notably, the same probability for all four if they are taken exactly as they are depicted. If taken as representatives of classes of similar configurations, however, their (subjective) probabilities are indeed inversely correlated to the conditional complexities for these individual configurations. For Bayesian inference, this implies that one could just as well use precisals instead of actual probabilities, because the decisive conditionals yield about the same predictive power in both cases (van der Helm, 2000).

Second, FE predictive coding seems to give action priority over perception, or as Friston (2009) put it: “*perception is an inevitable consequence of active exchange with the environment*” (p. 293) and “*perception is enslaved by action to provide veridical predictions*” (p. 295). However, as shown above, the role of action in everyday perception (or “active inference”, as Friston calls it)—though certainly relevant—is rather simple and straightforward. The foregoing also shows that the inclusion of action into the equation is not helpful in assessing what the first priors in perceptual organization might be. After all, as long as one has approximately the right conditionals, Bayesian inference works quite well for a moving observer—no matter which first

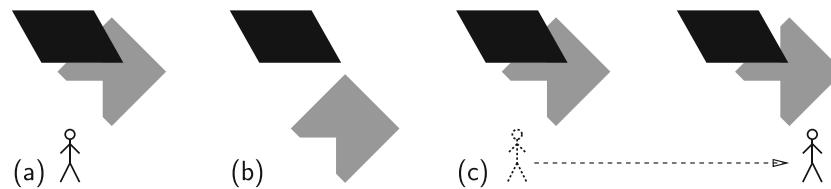


Fig. 3 Everyday perception by moving observers **a** You take a first glance at a scene **b** You probably interpret it as a black shape occluding this grey shape **c** You move, and what you see then may trigger a visual update leading to a revision of your first interpretation

priors are used. Yet, the question of the first priors is definitely relevant in human perception research, which, for instance, also aims to understand the perception of static images (which, in this multimedia era, probably are more abundantly present than in the past). Whereas FE predictive coding is silent about what the first priors might be (see also Footnote 4 and section “[Empirical priors](#)”), structural coding gives a principled answer by taking the precisal of a hypothesis as its first prior.

Discussion

As Hoffman (1996) put it in Bayesian terms: We have direct access to only the posteriors of perception. Hence, to understand these posteriors, we have to trace back what the priors and conditionals might have been. Bayes’ rule captures the interplay between priors and conditionals but does, of itself, not supply any specification of priors and conditionals. Therefore, standard Bayesian modeling involves model fitting to tune the parameters of a selection model such that it yields desired outcomes (this stands apart from hypothesis selection, i.e., the subsequent application of such a selection model to find hypotheses that meet the employed selection criterion). This powerful modeling method may well reflect learning strategies at higher cognitive levels, but in my view, perception plays a special role, which is to be distinguished from that of higher cognitive faculties.

Perception is sort of a communication channel, or interface, between the world and higher cognitive faculties. Following Leonardo da Vinci’s (1452–1519) motto “*All knowledge has its origins in perception*”, structural coding therefore takes perception as a fairly autonomous, data-driven, source of knowledge instead of taking knowledge as a resource for perception (cf. Firestone & Scholl, *in press*; Gottschaldt, 1926; Hochberg, 1978; Kanizsa, 1985; Pylyshyn, 1999; Rock, 1985). Furthermore, as said, the structural coding model aims to minimize the number of parameters needed to describe hypotheses (this is a form of hypothesis selection), but the structural coding model itself is basically parameter-free (so, no tuning of the selection model to get desired outcomes). In other words, by its simplicity principle, it gives a principled account of

priors and conditionals, which, as indicated, provides fairly optimal encoding of data and fairly veridical perception in daily life.

In perceptual organization, the Bayesian distinction between view-independent priors and view-dependent conditionals (be they precisals or other probabilities) concurs with the distinction between the ventral and dorsal streams in the brain, which seem to be dedicated to object perception and spatial perception, respectively (Ungerleider & Mishkin, 1982). The Bayesian integration of priors and conditionals can thus be said to model the interaction between these streams, which leads the visual system from percepts of objects as such to percepts of objects arranged in space. To structural coding, this is the (obviously grey) area where perception tends to end and higher cognitive faculties get the opportunity to enrich its output via a gradually more conscious inference on the basis of internally available contextual information (say, knowledge).

For instance, disks with shadings at the left-hand or right-hand side give fairly ambiguous impressions of concavity and convexity (see Fig. 4a), whereas disks with shadings at the top or bottom give fairly clear impressions of concavity and convexity, respectively (see Fig. 4b). By structural coding, all such disks are perceptually ambiguous. Yet, in some cases, such ambiguities might be resolved at higher cognitive levels (Rock, 1985)—here, for instance, by the

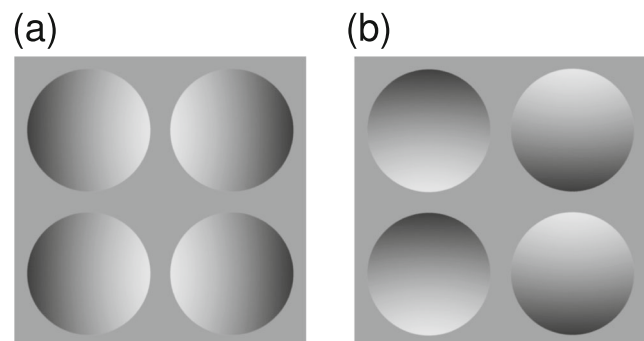


Fig. 4 Shape from shading. **a** Shading at the left-hand or right-hand side is fairly ambiguous regarding the concavity or convexity of the disks. **b** Shading at the top or bottom is fairly clear regarding concavity and convexity, respectively. (After Ramachandran, 1988)

knowledge that light usually comes from above. Some Bayesians incorporate such knowledge in models of perception but I do not think this is needed for the main task of perception, which is to organize incoming (meaningless) pieces of visual information into (meaningful) wholes and parts arranged in space.

This main task means that a percept reflects a hierarchical organization of a scene. In structural coding, candidate percepts (i.e., hypotheses) are assumed to be constructed from the sensory data and are represented by hierarchical codes, which impose such hierarchical organizations on the data. This contrasts with Bayesian approaches (including FE predictive coding), which are strong in capturing the interplay between probabilities of given hypotheses but which usually are silent about how these hypotheses are structured and represented (be it formally or in the brain).

In sum, regarding competence, FE predictive coding admittedly uses a powerful modeling technique, but in my view, structural coding has more explanatory power because of its principled account of priors and conditionals in terms of fairly stable descriptive complexities. It is also true, however, that FE predictive coding's main claims pertain not so much to competence but rather to performance. This is discussed next.

Performance

Traditional ideas about the human visual perceptual organization process have taken it to be nothing but a unidirectional, feedforward, process from sensory inputs to percepts. This holds neither for FE predictive coding nor for structural coding: Both coding approaches rely on recurrent and horizontal processing too. However, they put forward different forms of message passing. To compare them, I take Lee and Mumford's (2003) description of hierarchical Bayesian inference in the visual cortex as reference.

Hierarchical Bayesian inference

Lee and Mumford (2003) proposed a Bayesian predictive coding model that is not based on minimization of prediction errors. Instead, it takes visual area V1—which, via the lateral geniculate nucleus, receives input from the retina—as the first area to construct what they called particles, that is, preliminary interpretations of input parts. These particles are assumed to stay alive during a hierarchical inference process, by which a higher visual area takes particles from the previous area to construct its own larger particles, whose strength then is fed back to the previous area to allow for particle updating—and so on, until the system as a whole reaches an equilibrium. This process is called particle filtering, and during this process, particle updating is assumed

to be guided by Bayesian belief propagation. The latter means that the feedback from higher areas provides what they called contextual priors to shape the inference at lower areas.

Lee and Mumford allowed knowledge from higher cognitive levels (say, from beyond perception) to provide such feedback too, but notice that they essentially proposed a data-driven perceptual inference process, by which partial percepts (i.e., particles) interact and compete to arrive eventually at a complete percept. They were not specific about the internal representational structure of particles, but they did suggest that particles might be represented by temporarily synchronized neural assemblies.

Neuronal synchronization is the phenomenon that neurons, in transient assemblies, temporarily synchronize their firing activity. This is a special case of parallel distributed processing (PDP). That is, standard PDP typically involves interacting agents who simultaneously do different things, whereas synchronization involves interacting agents who simultaneously do the same thing—think of flash mobs or choirs going from cacophony to harmony. Both theoretically and empirically, neuronal synchronization has been associated with various cognitive processes and 30–70 Hz gamma-band synchronization, in particular, has been associated with feature binding in visual perceptual organization (Eckhorn et al., 1988; Gray & Singer, 1989; Milner 1974; von der Malsburg, 1981).

As I discuss next, FE predictive coding (which includes effects of knowledge) proposes hierarchical Bayesian inference too, but not in the form described by Lee and Mumford. As I discuss subsequently, structural coding (which excludes effects of knowledge) proposes a particle-filtering mechanism, but then with particle updating guided by propagation of the Occamian simplicity belief and with a computationally powerful specification of the representational role of neuronal synchronization in the gamma band.

FE predictive coding's cognitive architecture

Whereas Lee and Mumford's (2003) predictive coding approach holds that "*the feedforward input drives the generation of the hypotheses*" (p. 1436), FE predictive coding argues for more or less the reverse. It explicitly dismisses particle filtering (Friston, 2008, 2009) and relies instead on top-down testing of hypotheses against the sensory input (see Fig. 5). This top-down testing goes, in a hierarchical fashion, through the successive levels in the cortex—each level taken to be responsible for specific (intermediate) aspects. At each level, higher-level predictions are compared with lower-level sensory information to form a prediction error, which is returned to the higher level to enable it to update its predictions—these updated predictions then

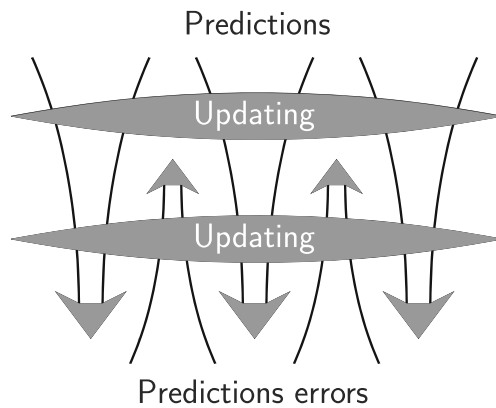


Fig. 5 FE predictive coding's view on processing in the brain's visual hierarchy. Predictions are tested top-down against the sensory input, and prediction errors are returned. At each level in the hierarchy, prediction errors are updated by combining messages from the same level and the level above, and predictions by combining messages from the same level and the level below

are recycled to reduce prediction errors at lower levels. In other words, feedforward connections convey information on prediction errors, while feedback connections convey information on predictions from higher cortical areas to suppress prediction errors in lower areas (Bastos et al., 2012).

Hence, FE predictive coding basically proposes a sort of glorified template matching. It is true that template matching can be effective in automatic recognition of things from a limited number of predefined categories, such as print characters or objects in an assembly line. However, in human vision research, it has been abandoned long ago because it is too rigid and limited to deal with ill-defined categories and novel objects. To be frank, I do not see how FE predictive coding's glorified version might turn the tables.

Empirical priors

As said, in FE predictive coding, feedback connections convey information on predictions from higher areas to suppress prediction errors in lower areas. This feedback is said to constitute empirical priors, which are claimed to dissolve the criticism of Bayesian models that they ignore the question of how prior beliefs—necessary for inference—are formed (Friston, 2010). However, notice that these empirical priors depend on the sensory data, that is, they actually are posteriors—which is not altered by the fact that they, just as in plain Bayesian inference (see section “[Bayesian inference and the role of action in perception](#)”), are fed back to become the new priors for the next inference cycle. In any case, they do not dissolve the just-mentioned criticism of Bayesian models, which is about first priors,

that is, about priors that are independent of the sensory data (see also Trappenberg & Hollensen, 2013). First priors are relevant, simply because they form the starting point of the inference process.

To be clear, the foregoing does not question feedback mechanisms as such. Feedback mechanisms are inherent to hierarchical inference models. For instance, as discussed in section “[Hierarchical Bayesian inference](#)”, Lee and Mumford's (2003) predictive coding model involves feedback of what they called contextual priors, that is, strengths of higher-level particles that had been composed of lower-level ones. Furthermore, structural coding did not invent a special name for the feedback information but it incorporates basically the same feedback mechanism as that in Lee and Mumford's model—except that it expresses particle strength in descriptive complexities instead of probabilities (see section “[Perception](#)”). In other words, I understand the relevance of the empirical priors in the FE predictive coding scheme, but I think that FE predictive coding gives them more credits than they deserve.

Attention and gamma synchronization

According to Friston (2009), attention simply is the process of optimizing the relative precision of feedforward and feedback information during the hierarchical inference process. Later, Friston (2010) and Bastos et al. (2012) suggested faintly that neuronal synchronization in the gamma band has something to do with prediction errors, and after that, Clark (2013) and Kanai et al. (2015) suggested that gamma synchronization controls the precision associated with prediction errors at lower levels relative to that at higher levels. Notice that, unlike what Friston (2009) attributed to attention, the latter applies to feedforward information only. Be that as it may, my present point is that there is no direct evidence, neither neurophysiologically nor otherwise, that attention or gamma synchronization controls the precision associated with prediction errors.

In other words, FE predictive coding's account rather seems to be a matter of reading into the facts, that is, of attempting to connect the favored approach to accepted phenomena—then, gamma synchronization might indeed be positionable only as being associated somehow with prediction errors. To be clear, I do not object to such attempts, but in this case, I think it is not convincing without, for instance, complementary formal support for the proposed computational role of gamma synchronization.

Structural coding's cognitive architecture

Compared to FE predictive coding, structural coding assumes other messages being passed up and down in the brain's visual hierarchy. Furthermore, structural coding

admittedly contains speculative components too, but it does supply complementary formal support for the computational role it attributes to gamma synchronization.

Structural coding's view on processing in the visual hierarchy includes both perception and (task-driven, top-down) attention. As discussed in van der Helm (2012, 2015a), it conceives of the perceptual organization process as comprising three neurally intertwined but functionally distinguishable subprocesses (see Fig. 6, left-hand panel; cf. Lamme & Roelfsema 2000; Lamme et al. 1998). These subprocesses are responsible for (a) feedforward extraction of, or tuning to, features to which the visual system is sensitive, (b) horizontal binding of similar features, and (c) recurrent selection of different features. These subprocesses together yield integrated percepts given by hierarchical organizations (i.e., organizations in terms of wholes and parts) of hypothesized distal stimuli that fit the sensory data (see Fig. 6, right-hand panel). Attentional processes then may scrutinize these organizations in a top-down fashion, that is, starting with global structures and, if required by task and allowed by time, descending to local features (Ahissar & Hochstein, 2004; Collard & Povel, 1982; Hochstein & Ahissar, 2002; Wolfe 2007). This may be specified further as follows for attention and perception, respectively.

Attention

Structural coding assumes that, guided by descriptive simplicity, the unconscious perception process arrives at complete percepts (i.e., perceived wholes) via nonlinear interactions between competing partial percepts (I return to

this in a moment). It assumes further a top-down attentional scrutiny of resulting hierarchical organizations, which implies that wholes are consciously experienced before parts. This explains the dominance of wholes over parts, as postulated in early twentieth century Gestalt psychology (Koffka, 1935; Köhler, 1920, 1929; Wertheimer, 1912, 1923) and as confirmed later in a range of behavioral studies (for a review, see Wagemans et al., 2012). This dominance means, for instance, that humans tend to classify or categorize things on the basis of their global structures (i.e., on the basis of wholes and ignoring minor differences in parts). Based on empirical data, it has been specified further by notions such as global precedence (Navon, 1977), configural superiority (Pomerantz, Sager, & Stoeber, 1977), primacy of holistic properties (Kimchi, 2003), and superstructure dominance (Leeuwenberg & van der Helm, 1991; Leeuwenberg, van der Helm, & van Lier, 1994).

To give an example of the dominance of wholes over parts, I consider Fig. 7. It shows a stimulus that is typically perceived as consisting of two triangular parts. These triangular parts therefore are said to be compatible with the perceived global structure, and they are more easily discerned than incompatible parts like the diamond in Fig. 7, bottom right. By Fig. 6, this can be understood as follows (see also van der Helm, 2015b). The perceptual organization process yields perceived hierarchical organizations in terms of global structures and their constituent local features. This means that it preserves the representations of the compatible constituents, and masks (or suppresses, or eliminates, or inhibits) those of incompatible parts. Thus, if a to-be-discerned local feature is compatible, the top-down attention

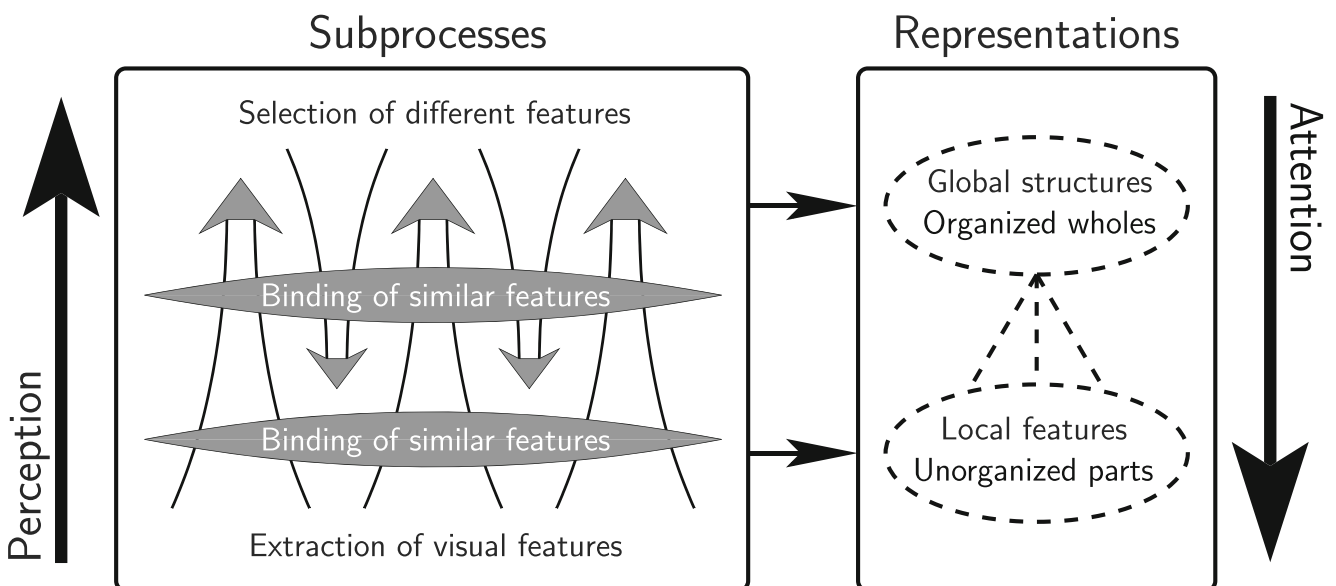


Fig. 6 Structural coding's view on processing in the brain's visual hierarchy. A stimulus-driven perceptual organization process, comprising three neurally intertwined subprocesses (*left-hand panel*), yields

hierarchical stimulus organizations (*right-hand panel*). A task-driven attention process then may scrutinize these hierarchical organizations in a top-down fashion

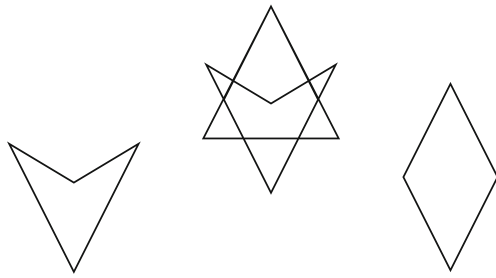


Fig. 7 Embedded figures. At the top, a stimulus with a typically perceived organization comprising two triangular shapes, plus one of these easily discerned compatible parts and an incompatible diamond part that is less easily discerned. (After Kastens & Ishikawa, 2006)

process may exploit the perceived hierarchical organization to descend easily from its global structure to this local feature. If it is not compatible—as is typical in embedded figures tasks, for instance—the top-down attention process first is misled by the perceived global structure and then has to find a way around it.

Perception

Among the perceptual subprocesses in Fig. 6, the subprocess of feedforward extraction is reminiscent of the neuroscientific idea that, going up in the visual hierarchy, neural cells mediate detection of increasingly complex features (Hubel & Wiesel, 1968). Furthermore, the subprocess of recurrent selection is reminiscent of the connectionist idea that a standard PDP process of activation spreading in the brain's neural network yields percepts represented by stable patterns of activation (Churchland, 1986). In structural coding, the combination of these two subprocesses is taken to be like a fountain under increasing water pressure: As the feedforward extraction progresses along ascending connections, each passed level in the visual hierarchy forms the starting point of integrative recurrent processing along descending connections. For a similar picture, see VanRullen and Thorpe (2002), and notice that this mechanism—just as Lee and Mumford's (2003) particle filtering—yields a gradual buildup from percepts of parts at lower levels in the visual hierarchy to percepts of wholes near its top end.

By nature, this gradual buildup takes time, so, it leaves room for attention to intrude and to modulate things before a percept has completed. In this sense, structural coding does not exclude influences from higher cognitive levels entirely. However, it also assumes that the perceptual organization process is very fast and that it, by then, already has done much of its integrative work (cf. Gray, 1999; Pylyshyn, 1999). Structural coding attributes this speed to neuronal synchronization in the gamma band. Notice that 30–70-Hz gamma oscillations are faster than 8–30-Hz alpha and beta oscillations. The latter usually are associated with top-down processes, while gamma synchronization occurs

predominantly in horizontal neural assemblies within visual areas, which have been associated with binding of similar features (Gilbert, 1992). The latter subprocess may be relatively underexposed in neuroscience, but may well be the neuronal counterpart of the regularity extraction operations, which, in representational coding approaches, are proposed to obtain structured mental representations of incoming visual information.

In fact, structural coding postulates that gamma synchronization mediates transparallel feature processing, which means that many similar features are hierarchically recoded in one go, that is, simultaneously as if only one feature were concerned (van der Helm, 2012, 2014, 2015a). There is no direct evidence that the brain indeed performs transparallel processing, but to my knowledge, it is the first computational proposal to do justice to the idea that neuronal synchronization must be a special form of neuro-cognitive processing. Moreover, this computational proposal is substantiated formally as follows.

Transparallel processing

In computing, transparallel processing corresponds to the extraordinarily powerful form of processing promised by quantum computers (see van der Helm, 2015a). Actually, it is already feasible on single-processor classical computers, and structural coding implemented it in PISA, which is a minimal coding algorithm for strings (van der Helm, 2004, 2015a). Notably, to compute guaranteed simplest codes of strings, PISA employs formal counterparts of the three perceptual subprocesses in Fig. 6. Through exploiting visual regularities such as repetition and symmetry, such codes specify strings by a minimum number of descriptive parameters (for formal and empirical underpinnings of the choice of employed regularities, see van der Helm & Leeuwenberg, 1991, 1996, 1999, 2004). Notice that a string gives rise to a superexponential number of candidate codes (i.e., hypotheses), so that the simplest one is probably not tractable by traditional forms of processing.

In PISA, this problem has been solved by employing special, usually sparse, distributed representations called hyperstrings. Hyperstrings are superpositions of up to an exponential number of similar regularities, which can be hierarchically recoded in a transparallel fashion, that is, simultaneously as if only one regularity were concerned.⁷

⁷The first version of PISA (van der Helm & Leeuwenberg, 1986) already employed what Lee and Mumford (2003) called particle filtering, but then, with particle updating guided by propagation of the Occamian simplicity belief. The second version (van der Helm, 1988) added the basics of what van der Helm (2004) formalized as transparallel processing by hyperstrings. The latter is fully implemented in the latest version, which is available in the Supplemental Material of van der Helm (2015a).

For more details, see van der Helm (2012, 2014, 2015a), in which hyperstrings are taken as formal counterparts of transient neural assemblies, while transparallel processing is proposed to be the special form of neuro-cognitive processing mediated by synchronization in such neural assemblies. Notice that this is consistent with Lee and Mumford's (2003) suggestion that particles might be represented by temporarily synchronized neural assemblies.

Strings do not, of course, constitute input like that of the human visual system. Nevertheless, the foregoing provides formal support for the idea that transparallel processing—mediated by gamma synchronization—might be the powerful form of neuro-cognitive processing needed to solve the superexponential inverse problem of perception.

Discussion

Neurophysiological evidence, on the one hand, links experimental conditions to brain activity but does, of itself, not indicate what this brain activity means in terms of cognitive information processing. Behavioral evidence, on the other hand, links experimental conditions to the outcome of cognitive information processing but does, of itself, not indicate how this outcome is arrived at. To dig deeper, one has to resort to performance models, or cognitive architectures as they are called in artificial intelligence research (Anderson, 1983; Newell, 1990). The architectures proposed by FE predictive coding and structural coding (see Figs. 5 and 6, respectively) are examples of such performance models. Clearly, both architectures still have to be elaborated further. Yet, it seems safe to say that FE predictive coding is ahead in its account of the neurophysiological side (see Clark, 2013), while structural coding is ahead regarding critical tests at the behavioral side (see Leeuwenberg & van der Helm, 2013) and regarding formal support for the proposed computational role of gamma synchronization (see van der Helm, 2012, 2014, 2015a).

Clarity about the role of gamma synchronization is particularly relevant to understand effects of impaired gamma synchronization, as found in neurodevelopmental disorders such as schizophrenia (Uhlhaas, Silverstein, & Phillips, 2005) and autism spectrum disorders (ASD) (Grice et al., 2001; Maxwell et al., 2015; Milne et al., 2009; Sun et al., 2012; Wright et al., 2012). For instance, within FE predictive coding, Clark (2013) suggested that impaired gamma synchronization leads to an imbalanced precision associated with prediction errors, that is, a higher precision at lower levels relative to that at higher levels (see also Kanai et al., 2015). Clark (2013) argued that this might explain hallucinations and delusions in schizophrenia (cf. Fletcher & Frith, 2009), but see also Silverstein (2013) who argued that these symptoms are more likely to arise at higher cognitive levels.

Furthermore, Lawson, Rees, and Friston (2014) argued that the imbalanced precision by impaired gamma synchronization idea explains various perceptual and social-exchange symptoms in ASD. However, without referring to gamma synchronization, van de Cruys et al. (2014) argued that those symptoms can be explained by a high inflexible precision of prediction errors at both lower and higher levels. As a consequence, van de Cruys et al. argued, ASD individuals put more value on small errors than typical individuals do. Whatever the relation between precision and gamma synchronization may be, putting more value on small errors agrees with findings that ASD individuals tend to focus more on local information in visual stimuli than on global information. For instance, they tend to categorize things into smaller categories (see, e.g., Klinger & Dawson, 2001; Newell et al., 2010) and are better in discerning embedded figures like the diamond in Fig. 7 (Jolliffe & Baron-Cohen, 1997; Shah & Frith, 1983, 1993).

In structural coding, the proposed role of gamma synchronization has nothing to do with prediction errors or their precision and is not, as in FE predictive coding, associated post-hoc with empirical data. It is based on formal computational grounds and implies that gamma synchronization subserves integration of local features into global structures (see section “[Transparallel processing](#)”). By this account, impaired gamma synchronization leads to less developed global structures. Such a reduced perceptual integration would affect classification abilities, and thereby, generalization and learning abilities (as seems to be the case in schizophrenia; see Doody et al., 1998). By the same token, it would result in categorization into smaller categories. Furthermore, linking up with Section “[Attention](#)”, it would also result in weaker masking effects on embedded figures (i.e., local features that are incompatible with typically perceived global structures), which therefore would be better discernable (van der Helm, 2015b). The latter agrees with the weak central coherence theory of ASD (Frith, 1989; see also Happé & Booth, 2008).

In other words, structural coding holds that, depending on the severity of the disorder, ASD individuals are left with something between incoming pieces of visual information and typically perceived wholes (think of an unfinished jigsaw puzzle). Then, top-down attention hardly has anything global to focus on, so, it naturally exhibits a narrowed focus and its access to embedded figures is hindered less by global structures. As van der Helm (2015b) argued, this also means that structural coding predicts that typical individuals are not worse than ASD individuals in discerning parts that are compatible with typically perceived global structures (like the triangular parts in Fig. 7)—simply because, in typical individuals, compatible parts are not masked by perceived global structures (see section “[Attention](#)”). As far

as I can tell, this is not what FE predictive coding would predict, so, future tests of this prediction may prove to be critical.

General discussion

FE predictive coding and structural coding both use free-energy minimization as metaphor for processing in the brain, but their elaborations of this metaphor are fundamentally different. FE predictive coding relies on classical information theory to minimize prediction errors, using probabilities to be tuned via model fitting. Structural coding relies on modern information theory to minimize the information load of predictions using fairly stable descriptive complexities. I am admittedly biased towards structural coding, but in this article, I have tried to make a fair assessment of FE predictive coding.

To be frank, I found it hard to deconstruct. For instance, in section “[FE predictive coding’s cognitive architecture](#)”, I indicated that template matching has been abandoned long ago in human vision research and that I do not see how FE predictive coding’s glorified version might turn the tables. Furthermore, its sometimes grandiloquent statements often seem to capitalize on intuitive associations in readers. One example thereof is its usage of the association-laden term surprise instead of the formal term surprisal from classical information theory. Another example is Bastos et al.’s (2012) “*through selecting appropriate sensations, the brain is implicitly maximizing the evidence for its own existence*” (p. 702; see also Friston, 2010). To me, the last part is esoterism, and I would not say that the brain selects appropriate sensations. I simply would say instead that, through action, it can select different vantage points but will have to do with whatever sensations it gets. In this sense, as indicated in section “[Bayesian inference and the role of action in perception](#)”, I think that FE predictive coding exaggerates the role of action in perception.

Be that as it may, notice that I sympathize with the more general Bayesian brain idea—albeit that I make a clear functional distinction between perception and higher cognitive levels. For instance, I can appreciate that—to increase practical utility—one might want to include knowledge (e.g., about the environment) in machine vision systems. However, I think that—due to transparallel processing—the human perceptual organization process is so fast that it hardly leaves room for effects of such knowledge, and that such apparent effects rather reflect post-perceptual enrichment. I therefore think that knowledge-based Bayesian approaches might be suited to model inferences at higher cognitive levels, but that perceptual inferences rather are guided by the Occamian simplicity belief working on data to construct, on the fly, hypotheses about these data.

Structural coding pursues the latter, accomplishing much of what FE predictive coding aims to accomplish—including links from perception to attention and action. Structural coding needs further elaboration, particularly at the neurophysiological side. Yet, as discussed, it is basically a parameter-free approach, which, by its simplicity principle, gives a principled account of priors and conditionals, providing fairly optimal encoding of data and fairly veridical perception in daily life. To this end, it relies on minimal coding of the internal structure of individual messages, which seems an appropriate reflection of the way in which the brain might encode sensory information in an efficient and parsimonious fashion. Furthermore, it substantiates that transparallel processing—mediated by gamma synchronization—might be the form of neuro-cognitive processing that solves the inverse problem of perception by way of a flexible, self-organizing, cognitive architecture implemented in the relatively rigid neural architecture of the brain.

In structural coding’s minimal-coding algorithm PISA, transparallel processing is enabled by hyperstrings, which are distributed representations built on the fly by the subprocess of horizontal feature binding and operated on by the subprocess of recurrent feature selection. By these data structures, structural coding links up with network models (see van der Helm, 2012). Furthermore, by converting descriptive complexities into precisals, structural coding might be given a Bayesian formulation that, in various respects, would resemble Lee and Mumford’s (2003) model. In other words, structural coding can be said to represent a separate branch of the diverse family of predictive coding models.

Finally, my critique of FE predictive coding should not obscure that I do appreciate that it—just as other predictive coding approaches and just as structural coding—aims to unify ideas about competence and performance. The distinction between these two notions corresponds to the distinction between what Wertheimer called the molar and molecular levels (see Koffka, 1935) or what Marr (1982/2010) called the “what” and “how” questions. As Marr noted, answering these questions may be totally different endeavors, but answers to both questions are needed for a full understanding.

Notes I thank Sander Van de Cruys for helpful discussions. This research was supported by Methusalem grant METH/08/02 awarded to Johan Wagemans (www.gestaltrevision.be).

References

- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Science*, 8, 457–464. doi:10.1016/j.tics.2004.08.011

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193. doi:10.1037/h0054663
- Austerweil, J. L., & Griffiths, T. L. (2013). Constructing flexible feature representations using nonparametric Bayesian inference. *Psychological Review*, *120*, 817–851. doi:10.1037/a0034194
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*, 695–711. doi:10.1016/j.neuron.2012.10.038
- Bayes, T., & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, *53*, 370–418. doi:10.1098/rstl.1763.0053
- Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*, 566–581. doi:10.1037/0033-295X.103.3.566
- Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204. doi:10.1017/S0140525X12000477
- Collard, R. F., & Povel, D. J. (1982). Theory of serial pattern production: Tree traversals. *Psychological Review*, *89*, 693–707. doi:10.1037/0033-295X.89.6.693
- Diaconis, P., & Freedman, D. (1986). On the consistency of Bayes estimates. *Annals of Statistics*, *14*, 1–26. doi:10.1214/aos/1176349830
- Doody, G. A., Johnstone, E. C., Sanderson, T. L., Owens, D. G. C., & Muir, W. J. (1998). "Pfpofschizophrenie" revisited: Schizophrenia in people with mild learning disability. *British Journal of Psychiatry*, *173*, 145–153. doi:10.1192/bjp.173.2.145
- Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., & Reitboeck, H. J. (1988). Coherent oscillations: A mechanisms of feature linking in the visual cortex?. *Biological Cybernetics*, *60*, 121–130. doi:10.1007/BF00202899
- Feldman, J. (2007). Formation of visual "objects" in the early computation of spatial relations. *Perception & Psychophysics*, *69*, 816–827. doi:10.3758/BF03193781
- Feldman, J. (2009). Bayes and the simplicity principle in perception. *Psychological Review*, *116*, 875–887. doi:10.1037/a0017144
- Feldman, J. (2013). Tuning your priors to the world. *Topics in Cognitive Science*, *5*, 13–34. doi:10.1111/tops.12003
- Firestone, C., & Scholl, B. (in press). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*.
- Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*, 48–58. doi:10.1038/nrn2536
- Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*, e1000211. doi:10.1371/journal.pcbi.1000211
- Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in Cognitive Science*, *13*, 293–301. doi:10.1016/j.tics.2009.04.005
- Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience*, *11*, 127–138. doi:10.1038/nrn2787
- Frith, U. (1989). *Autism: Explaining the enigma*. Oxford, UK: Basil Blackwell.
- Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York, NY: Wiley.
- Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron*, *9*, 1–13. doi:10.1016/0896-6273(92)90215-Y
- Gottschaldt, K. (1926). Ueber den Einfluss der Erfahrung auf die Wahrnehmung von Figuren [On the influence of experience on the perception of form]. *Psychologischen Forschungen*, *8*, 261–317. doi:10.1007/BF02411523
- Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: Still alive and well. *Neuron*, *24*, 31–47. doi:10.1016/S0896-6273(00)80820-X
- Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences USA*, *86*, 1698–1702. doi:10.1073/pnas.86.5.1698
- Gregory, R. L. (1973). The confounded eye. In Gregory, R., & Gombrich, E. (Eds.) *Illusion in nature and art* (pp. 49–95). London: Duckworth.
- Grice, S. J., Spratling, M. W., Karmiloff-Smith, A., Halit, H., Csibra, G., De Haan, M., & Johnson, M. H. (2001). Disordered visual processing and oscillatory brain activity in autism and Williams Syndrome. *NeuroReport*, *12*, 2697–2700. doi:10.1097/00001756-200108280-00021
- Happé, F. G. E., & Booth, R. D. L. (2008). The power of the positive: Revisiting weak coherence in autism spectrum disorders. *The Quarterly Journal of Experimental Psychology*, *61*, 50–63. doi:10.1080/17470210701508731
- Hochberg, J. E. (1978). *Perception*, 2nd. Englewood Cliffs, NJ: Prentice-Hall.
- Hochberg, J. E., & McAlister, E. (1953). A quantitative approach to figural "goodness". *Journal of Experimental Psychology*, *46*, 361–364. doi:10.1037/h0055809
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791–804. doi:10.1016/S0896-6273(02)01091-7
- Hoffman, D. D. (1996). What do we mean by "The structure of the world"? In Knill, D. C., & Richards, W. (Eds.) *Perception as Bayesian Inference* (pp. 219–221). Cambridge, MA: Cambridge University Press.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, *195*, 215–243. doi:jphysiol.1968.sp008455
- Jolliffe, T., & Baron-Cohen, S. J. (1997). Are people with autism and Asperger syndrome faster than normal on the Embedded Figures Test?. *Journal of Child Psychology and Psychiatry*, *38*, 527–534. doi:10.1111/j.1469-7610.1997.tb01539.x
- Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B*, *370*, 20140169. doi:10.1098/rstb.2014.0169
- Kanizsa, G. (1985). Seeing and thinking. *Acta Psychologica*, *59*, 23–33. doi:10.1016/0001-6918(85)90040-X
- Kastens, K. A., & Ishikawa, T. (2006). Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields. *Geological Society of America Special Papers*, *413*, 53–76. doi:10.1130/2006.2413(05)
- Kimchi, R. (2003). Relative dominance of holistic and component properties in the perceptual organization of visual objects. In Peterson, M. A., & Rhodes, G. (Eds.) *Perception of faces, objects, and scenes: Analytic and holistic processes*. doi:10.1093/acprof:oso/9780195313659.003.0010 (pp. 235–263). New York, NY: Oxford University Press.
- Klinger, L., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology*, *13*, 111–124. doi:10.1017/S0954579401001080
- Koffka, K. (1935). *Principles of Gestalt psychology*. London: Routledge & Kegan Paul.
- Köhler, W. (1920). *Die physischen Gestalten in Ruhe und im stationären Zustand [Static and stationary physical shapes]*. Braunschweig, Germany: Vieweg.
- Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems in Information Transmission*, *1*, 1–7. doi:10.1080/00207166808803030

- Lamme, V.A.F., & Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, 23, 571–579. doi:10.1016/S0166-2236(00)01657-X
- Lamme, V.A.F., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, 8, 529–535. doi:10.1016/S0959-4388(98)80042-1
- Lawson, R.P., Rees, G., & Friston, K.J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, 8, 302. doi:10.3389/fnhum.2014.00302
- Lee, T.S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20, 1434–1448. doi:10.1364/JOSAA.20.001434
- Leeuwenberg, E.L.J. (1968). *Structural information of visual patterns: an efficient coding system in perception*. The Hague-Paris: Mouton & Co.
- Leeuwenberg, E.L.J., & van der Helm, P.A. (1991). Unity and variety in visual form. *Perception*, 20, 595–622. doi:10.1068/p200595
- Leeuwenberg, E.L.J., & van der Helm, P.A. (2013). *Structural information theory: The simplicity of visual form*. Cambridge, UK: Cambridge University Press.
- Leeuwenberg, E.L.J., van der Helm, P.A., & van Lier, R.J. (1994). From geons to structure: A note on object classification. *Perception*, 23, 505–515. doi:10.1068/p230505
- Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications*, 2nd. New York: Springer-Verlag.
- MacKay, D. (1950). Quantal aspects of scientific information. *Philosophical Magazine*, 41, 289–311. doi:10.1080/14786445008521798
- MacKay, D.J.C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK: Cambridge University Press.
- Marr, D. (2010). *Vision*. Cambridge, MA: MIT Press. (Original work published 1982 by Freeman).
- Maxwell, C.R., Villalobos, M.E., Schultz, R.T., Herpertz-Dahlmann, B., Konrad, K., & Kohls, G. (2015). Atypical laterality of resting gamma oscillations in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 45, 292–297. doi:10.1007/s10803-013-1842-7
- Milne, E., Scope, A., Pascalis, O., Buckley, D., & Makeig, S. (2009). Independent component analysis reveals atypical electroencephalographic activity during visual perception in individuals with autism. *Biological Psychiatry*, 65, 22–30. doi:10.1016/j.biopsych.2008.07.017
- Milner, P. (1974). A model for visual shape recognition. *Psychological Review*, 81, 521–535. doi:10.1037/h0037149
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383. doi:10.1016/0010-0285(77)90012-3
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Newell, L.C., Best, C.A., Gastgeb, H., Rump, K.M., & Strauss, M.S. (2010). The development of categorization and facial knowledge: Implications for the study of autism. In Oakes, L.M., Cashon, C.H., Casasola, M., & Rakison, R.H. (Eds.) *Infant perception and cognition: Recent advances, emerging theories, and future directions (Chapter 11)*. doi:10.1093/acprof:oso/9780195366709.003.0011 Oxford, UK: Oxford University Press.
- Pomerantz, J., & Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In Boff, K.R., Kaufman, L., & Thomas, J.P. (Eds.) *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 36-1–36-46). New York: Wiley.
- Pomerantz, J.R., Sager, L.C., & Stoeber, R.J. (1977). Perception of wholes and their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 422–435. doi:10.1037/0096-1523.3.3.422
- Pylyshyn, Z.W. (1999). Is vision continuous with cognition? The case of impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341–423. doi:10.1017/S0140525X99002022
- Ramachandran, V.S. (1988). Perception of shape from shading. *Nature*, 331, 163–166. doi:10.1038/331163a0
- Rao, R.P.N., & Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience*, 2, 79–87. doi:10.1038/4580
- Rissanen, J.J. (1978). Modelling by the shortest data description. *Automatica*, 14, 465–471. doi:10.1016/0005-1098(78)90005-5
- Rock, I. (1985). Perception and knowledge. *Acta Psychologica*, 59, 3–22.
- Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry*, 24, 613–620. doi:10.1111/j.1469-7610.1983.tb00137.x
- Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the block design task?. *Journal of Child Psychology and Psychiatry*, 34, 1351–1364. doi:10.1111/j.1469-7610.1993.tb02095.x
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x
- Silverstein, S.M. (2013). Schizophrenia-related phenomena that challenge prediction error as the basis of cognitive functioning. *Behavioral and Brain Sciences*, 36, 229–230. doi:10.1017/S0140525X12002221
- Simon, H.A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 79, 369–382. doi:10.1037/h0033118
- Solomonoff, R.J. (1964a). A formal theory of inductive inference, Part 1. *Information and Control*, 7, 1–22. doi:10.1016/S0019-9958(64)90223-2
- Solomonoff, R.J. (1964b). A formal theory of inductive inference, Part 2. *Information and Control*, 7, 224–254. doi:10.1016/S0019-9958(64)90131-7
- Sun, L., Grützner, C., Bölte, S., Wibral, M., Tozman, T., Schlitt, S., Poustka, F., Singer, W., Freitag, C.M., & Uhlhaas, P.J. (2012). Impaired gamma-band activity during perceptual organization in adults with autism spectrum disorders: evidence for dysfunctional network activity in frontal-posterior cortices. *Journal of Neuroscience*, 32, 9563–9573. doi:10.1523/JNEUROSCI.1073-12.2012
- Thornton, C. (2014). Infotropism as the underlying principle of perceptual organization. *Journal of Mathematical Psychology*, 61, 38–44. doi:10.1016/j.jmp.2014.08.002
- Trappenberg, T., & Hollensen, P. (2013). Sparse coding and challenges for Bayesian models of the brain. *Behavioral and Brain Sciences*, 36, 232–233. doi:10.1017/S0140525X12002300
- Tribus, M. (1961). *Thermostatistics and thermodynamics*. Princeton, NJ: Van Nostrand.
- Uhlhaas, P.J., Silverstein, S.M., & Phillips, W.A. (2005). The course and clinical correlates of dysfunctions in visual perceptual organization in schizophrenia during the remission of psychotic symptoms. *Schizophrenia Research*, 75, 183–192. doi:10.1016/j.schres.2004.11.005
- Ungerleider, L.G., & Mishkin, M. (1982). Two cortical visual systems. In Ingle, D.J., Goodale, M.A., & Mansfield, R.J.W. (Eds.) *Analysis of Visual Behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., De-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, 121, 649–675. doi:10.1037/a0037665

- van der Helm, P. A. (1988). *Accessibility and simplicity of visual structures*. The Netherlands: Ph.D. thesis, Radboud University Nijmegen.
- van der Helm, P. A. (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, *126*, 770–800. doi:[10.1037//0033-2909.126.5.770](https://doi.org/10.1037//0033-2909.126.5.770)
- van der Helm, P. A. (2004). Transparallel processing by hyperstrings. *Proceedings of the National Academy of Sciences USA*, *101*(30), 10862–10867. doi:[10.1073/pnas.0403402101](https://doi.org/10.1073/pnas.0403402101)
- van der Helm, P. A. (2011). Bayesian confusions surrounding simplicity and likelihood in perceptual organization. *Acta Psychologica*, *138*, 337–346. doi:[10.1016/j.actpsy.2011.09.007](https://doi.org/10.1016/j.actpsy.2011.09.007)
- van der Helm, P. A. (2012). Cognitive architecture of perceptual organization: From neurons to gnosaurs. *Cognitive Processing*, *13*, 13–40. doi:[10.1007/s10339-011-0425-9](https://doi.org/10.1007/s10339-011-0425-9)
- van der Helm, P. A. (2014). *Simplicity in vision: A multidisciplinary account of perceptual organization*. Cambridge, UK: Cambridge University Press.
- van der Helm, P. A. (2015a). Transparallel mind: Classical computing with quantum power. *Artificial Intelligence Review*, *44*, 341–363. doi:[10.1007/s10462-015-9429-7](https://doi.org/10.1007/s10462-015-9429-7)
- van der Helm, P. A. (2015b). A cognitive architecture account of the visual local advantage phenomenon in autism spectrum disorders. *Vision Research (Online First)*. doi:[10.1016/j.visres.2015.04.009](https://doi.org/10.1016/j.visres.2015.04.009)
- van der Helm, P. A., & Leeuwenberg, E. L. J. (1986). Avoiding explosive search in automatic selection of simplest pattern codes. *Pattern Recognition*, *19*, 181–191. doi:[10.1016/0031-3203\(86\)90022-1](https://doi.org/10.1016/0031-3203(86)90022-1)
- van der Helm, P. A., & Leeuwenberg, E. L. J. (1991). Accessibility, a criterion for regularity and hierarchy in visual pattern codes. *Journal of Mathematical Psychology*, *35*, 151–213. doi:[10.1016/0022-2496\(91\)90025-O](https://doi.org/10.1016/0022-2496(91)90025-O)
- van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review*, *103*, 429–456. doi:[10.1037/0033-295X.103.3.429](https://doi.org/10.1037/0033-295X.103.3.429)
- van der Helm, P. A., & Leeuwenberg, E. L. J. (1999). A better approach to goodness: Reply to Wagemans (1999). *Psychological Review*, *106*, 622–630. doi:[10.1037/0033-295X.106.3.622](https://doi.org/10.1037/0033-295X.106.3.622)
- van der Helm, P. A., & Leeuwenberg, E. L. J. (2004). Holographic goodness is not that bad: Reply to Olivers, Chater, and Watson (2004). *Psychological Review*, *111*, 261–273. doi:[10.1037/0033-295X.111.1.261](https://doi.org/10.1037/0033-295X.111.1.261)
- van Lier, R. J., van der Helm, P. A., & Leeuwenberg, E. L. J. (1994). Integrating global and local aspects of visual occlusion. *Perception*, *23*, 883–903. doi:[10.1068/p230883](https://doi.org/10.1068/p230883)
- VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, *42*, 2593–2615. doi:[10.1016/S0042-6989\(02\)00298-5](https://doi.org/10.1016/S0042-6989(02)00298-5)
- von Malsburg, C. (1981). *The correlation theory of brain function. Internal Report 81-2, Max-Planck-Institute for Biophysical Chemistry*. Germany: Göttingen.
- von Helmholtz, H. L. F. (1962). *Treatise on physiological optics (J.P.C. Southall, Transactions)*. New York: Dover. (Original work published 1909).
- Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, *138*, 1218–1252. doi:[10.1037/a0029334](https://doi.org/10.1037/a0029334)
- Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung [Experimental study on the perception of movement]. *Zeitschrift für Psychologie*, *12*, 161–265.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II [On Gestalt theory]. *Psychologische Forschung*, *4*, 301–350. doi:[10.1007/BF00410640](https://doi.org/10.1007/BF00410640)
- Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In Gray, W. (Ed.) *Integrated models of cognitive systems* (pp. 99–119). New York: Oxford University Press.
- Wolpert, D. H. (1995). On the Bayesian "Occam factors" argument for Occam's razor. In Petsche, T., Hanson, S. J., & Shavlik, J. W. (Eds.) *Computational learning theory and natural learning systems, vol. III: Selecting good models (Chapter 11)*. Cambridge, MA: MIT Press.
- Wright, B., Alderson-Day, B., Prendergast, G., Bennett, S., Jordan, J., Whitton, C., Gouws, A., Jones, N., Attur, R., Tomlinson, H., & Green, G. (2012). Gamma activation in young people with autism spectrum disorders and typically-developing controls when viewing emotions on faces. *PLoS One*, *7*, e41326. doi:[10.1371/journal.pone.0041326](https://doi.org/10.1371/journal.pone.0041326)