

Simplicity in perceptual organization

Peter A. van der Helm
University of Leuven (KU Leuven), Belgium
peter.vanderhelm@ppw.kuleuven.be

To appear in

Oxford Handbook of Perceptual Organization
Oxford University Press
Edited by Johan Wagemans

Abstract

A long-standing debate in perception concerns the question of whether perceptual organization is guided by internal efficiency (the simplicity principle) or by external veridicality (the likelihood principle). This article focuses on the simplicity principle which is a modern information-theoretic version of Occam's razor, but also compares it to the likelihood principle which reflects classical information theory. Both principles can be modeled by Bayes' rule which combines quantifications of view-independent and view-dependent factors to predict stimulus interpretations. Whereas the likelihood principle relies on hardly quantifiable frequencies of occurrence of things in the world, the simplicity principle relies on better quantifiable structural complexities of individual things. The simplicity principle is further argued to be sufficiently veridical in everyday perception, and neurally realizable via cognitive architecture enabling transparallel processing. This quantum-like form of representational processing relates to connectionist modeling and complements dynamic-systems ideas about neuronal synchronization.

Keywords

perceptual organization; simplicity versus likelihood; classical versus modern information theory; Occam's razor; Bayesian modeling; view-independencies versus view-dependencies; veridicality; cognitive architecture; transparallel processing; neuronal synchronization

Acknowledgement

Preparation of this chapter was supported by Methusalem grant METH/08/02 awarded to Johan Wagemans (www.gestaltrevision.be)

1. Introduction

Perceptual organization is the neuro-cognitive process that takes the light in our eyes as input and that enables us to interpret scenes as structured wholes consisting of objects arranged in space - wholes which, moreover, usually are sufficiently veridical to guide action. This automatic process may seem to occur effortlessly, but by all accounts, it must be very complex and yet very flexible. To organize meaningless patches of light into meaningfully structured wholes within (literally) the blink of an eye, it must combine a high combinatorial capacity with a high speed (notice that a recognition model that tests previously stored templates against the visual input might avoid the combinatorics but would not achieve the required speed). To give a gist (following Gray, 1999, but many others have argued similarly), multiple sets of features at multiple, sometimes overlapping, locations in a stimulus must be grouped simultaneously. This implies that the process must cope with a large number of possible combinations in parallel, which also suggests that these possible combinations are engaged in a stimulus-dependent competition between grouping criteria. Hence, the combinatorial capacity of the perceptual organization process must be very high. This, together with its high speed (it completes in the range of 100-300 ms), reveals the truly impressive nature of the perceptual organization process.

One of the great mysteries of perception is how the human visual system manages to do all this. An intriguing idea in this context is that, from among all possible interpretations of a stimulus, the visual system selects the one defined by a minimum number of parameters. This simplicity principle has gained empirical support but is also controversial. Indeed, simplicity is obviously an appealing property in many settings, but can it be the guiding principle of the intricate process sketched above? To review this idea, this chapter focuses on underlying theoretical issues which may be introduced by way of a brief history of this principle.

2. A brief history of simplicity

An early predecessor of the simplicity principle is what became known as Occam's razor. Its origins can be traced back to Aristotle (384–322 BC), and it entails the advice - expressed in various forms by William of Occam (±1290–1349) - to keep theories and models as simple as possible, that is, to not make them more complex than needed to account for the available data. The underlying idea is that, all else being equal, the simplest of all possible interpretations of data is the best one. A modern version of Occam's razor is Rissanen's (1978) minimum description length principle (MDL principle) in the mathematical domain of algorithmic information theory (AIT, a.k.a. the theory of Kolmogorov complexity; Li & Vitányi, 1997). The MDL principle applies to model selection and, more

general, to inductive inference (Solomonoff, 1964a, 1964b). It proposes a trade-off between the complexity of hypotheses as such and their explanatory power, as follows:

The best hypothesis to explain given data is the one that minimizes the sum of

- (a) the information needed to describe the hypothesis; and
- (b) the information needed to describe the data with the help of the hypothesis.

For instance, in physics, Einstein's theory as such is more complex than that of Newton, but because it explains much more data, it is yet considered to be better. Applied to perceptual organization, the two amounts of information above can be taken to refer to, respectively, the view-independent complexity of hypothesized distal stimuli as such and their view-dependent degree of consistency with the proximal stimulus at hand. The MDL principle then suggests that, in the absence of further knowledge, the best interpretation of a stimulus is the one that minimizes the sum of these two amounts of information.

Another predecessor of the simplicity principle is the law of Prägnanz. The early 20-th century Gestalt psychologists Wertheimer (1912, 1923), Köhler (1920), and Koffka (1935) proposed that this law underlies perceptual groupings based on properties such as symmetry and similarity. It was inspired by the minimum principle in physics, which holds that dynamic physical systems tend to settle into relatively stable states defined by minimum energy loads. Applied to perceptual organization, the law of Prägnanz suggests that, when faced with a stimulus, the human visual system tends to settle into relatively stable neural states reflecting cognitive properties such as symmetry and simplicity. This idea does not exclude the influence of knowledge represented at higher cognitive levels, but it takes this influence to be subordinate to stimulus-driven mechanisms of a largely autonomous visual system.

Nowadays, the neural side of the law of Prägnanz finds elaboration in connectionist and dynamic-systems approaches to cognition. In the spirit of Marr's (1982/2010) levels of description, these two kinds of approaches are complementary in that connectionism usually focuses on the internal mechanisms of information processing systems, while dynamic systems theory (DST) usually focuses on the physical development over time of whole systems. Also complementary, but then usually focusing on the nature of outcomes of information processes, is representational theory in which the cognitive side of the law of Prägnanz finds elaboration. This may be specified as follows. For perceptual organization, Koffka (1935) formulated the law of Prägnanz as holding:

of several geometrically possible organizations that one will actually occur

which possesses the best, the most stable shape (p. 138),

and Hochberg and McAlister (1953) put this in information-theoretic terms by:

the less the amount of information needed to define a given organization as compared to the other alternatives, the more likely that the figure will be so perceived (p. 361),

specifying descriptive information loads, or complexities, by:

the number of different items we must be given, in order to specify or reproduce a given pattern (p. 361).

Hochberg and McAlister coined this information-theoretic idea the descriptive minimum principle, and nowadays, it is also known as the simplicity principle.

Hence, just as the MDL principle in AIT, the simplicity principle in perception promotes simplest codes as specifying the outcomes of an inference process based on descriptive codes of things. Such descriptive codes are much like computer codes, that is, representations that can be seen as reproduction recipes for things and whose internal structures are therefore enforced by the internal structures of those things. Both the MDL principle and the simplicity principle reflect modern information-theoretic approaches which contrast with Shannon's (1948) classical selective-information approach in communication theory. Shannon's approach promotes optimal codes, that is, nominalistic label codes (as in the Morse code) that minimize the long-time average burden on communication channels - assuming the transmission probabilities of codes are known. The simplicity principle further contrasts with von Helmholtz' (1909/1962) likelihood principle. The latter holds that the internal neuro-cognitive process of perceptual organization is guided by veridicality and yields interpretations most likely to be true in the external world - assuming such probabilities are known. Shannon's and von Helmholtz' approaches are appealing but suffer from the problem that, in many situations, the required probabilities are unknown if not unknowable. A main objective of modern descriptive-information theory is to circumvent this problem, that is, to make inferences without having to know the real probabilities.

An initial problem for modern information theory was that complexities depend on the chosen descriptive coding language. However, both theoretical findings in AIT (Chaitin, 1969; Kolmogorov, 1965; Solomonoff, 1964a, 1964b) and empirical findings in perception (Simon, 1972) provided evidence that, regarding complexity rankings, it does not matter much which descriptive coding language is employed. This evidence is not solid proof, but does suggest that descriptive simplicity is a fairly stable

concept.

The simplicity principle in perception agrees with ideas by Attneave (1954, 1982) and Garner (1962, 1974), for instance, and it has been promoted most prominently in Leeuwenberg's (1968, 1969, 1971) structural information theory (SIT). SIT was developed independently of AIT, but in hindsight, its current implementation of the simplicity principle can be seen as a perception-tailored version of the MDL principle in AIT (van der Helm, 2000). A notable difference, though, is that the MDL principle postulates that simplest interpretations are the best ones (without qualifying what "best" means), whereas the simplicity principle postulates that they are the ones most likely to result from the internal neuro-cognitive process of perceptual organization - which may not be interpretations most likely to be true in the external world.

This historical overview raises three questions which, below, are discussed in more detail. The first question is whether the human visual system indeed organizes stimuli in the simplest way; this is basically an empirical question, but because it has been plagued by unclarities, it is addressed by looking at operationalizations of simplicity. The second question is whether simplest stimulus organizations are sufficiently veridical; this is a theoretical question which is addressed by using AIT findings in a comparison between the simplicity and likelihood principles. The third question is whether the simplicity principle agrees with the putative high combinatorial capacity and speed of perceptual organization; this is a tractability question which is addressed by relating SIT to DST and connectionism to assess how the simplicity principle might be neurally realized.

3. Operationalizations of simplicity

Hochberg and McAlister (1953) introduced the simplicity principle in an article entitled *A quantitative approach to figural "goodness"*. Figural goodness is an intuitive Gestalt notion and the idea behind the association between descriptive simplicity and goodness is that simplicity entails both accuracy and parsimony. For instance, a square can be represented as if it were a rectangle, but representing it as a square is both more accurate and more efficient in terms of memory resources as it requires fewer descriptive parameters. Assuming that patterns are represented in the simplest way, simpler patterns are thus expected to be better in the sense that they can be remembered or reproduced more easily.

Hence, the motto here is "what is simple, is easy to learn". Notice that this is the inverse of the motto "what has been learned, is simple" which expresses that patterns, that have been seen often, are familiar so that they are experienced as being simple. The latter motto agrees with the

likelihood principle rather than with the simplicity principle, but it shows that simplicity has different connotations which may be relevant in different settings (see also Sober, 2002). Therefore, this section first addresses this issue.

3.1. Classical versus modern information-theoretic simplicity

In classical selective-information theory, the idea is that simpler things are things that convey less information because they belong to larger sets of actually occurring equivalent things (i.e., identical things, or similar things if their dissimilarities can be ignored in the situation at hand). A random dot cloud, for instance, is thus said to be simple: the set of random dot clouds is larger than any set of more structured dot patterns, so that a randomly picked dot pattern has a relatively high probability of being a random dot cloud. It therefore gets a shorter optimal code in Shannon's selective-information approach.

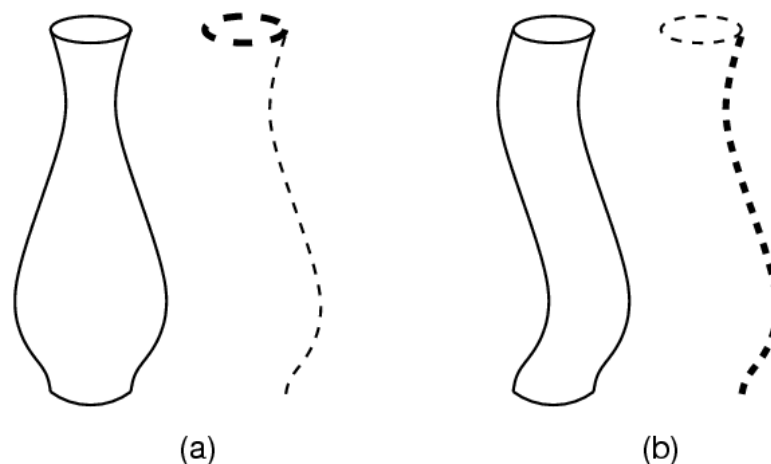


Figure 1. Objects that are simple because they have a highly regular internal structure consisting of a superstructure (visualized by thick dashes) that determines the positions of many identical subordinate structures (visualized by thin dashes). The hierarchy in (a) is the inverse of that in (b), and in both cases, the objects are presumably classified on the basis of primarily the perceptually dominant superstructure.

The objects in Figure 1, on the other hand, can be said to be simple in the sense that they have a highly regular internal structure. This idea about simplicity agrees with modern descriptive-information theory, in which individual things get shorter descriptive codes if they contain more structural regularity. This time, things may also be simple for another reason, by the way. For instance, the binary string 1111111111 is simple because it contains a structural regularity as all bits are identical, while the binary string 01 is simple because it contains only two bits. Shortest descriptive codes account for the simplicity of both cases, but for the rest, the two cases are hardly comparable. This illustrates that the

complexity of simplest codes is not always the most appropriate property to be used in inter-stimulus comparisons (i.e., in comparisons between interpretations of different stimuli). Indeed, the simplicity principle applies primarily to intra-stimulus comparisons, that is, to comparisons between different candidate interpretations of an individual stimulus. Furthermore, beside the complexity, also other properties of simplest codes may be used in inter-stimulus comparisons. For instance, unlike optimal codes, simplest codes have a hierarchical structure reflecting the hierarchical structure of simplest stimulus organizations, so that classifications of different stimuli may be assessed on the basis of these hierarchical code structures (see Figure 1; for more examples, see Leeuwenberg & van der Helm, 2013).

These different ideas about simplicity are also reflected in the following. In classical information theory, the length of an optimal code for an individual pattern is determined by the size of the set of all actually occurring identical patterns. In modern information theory, conversely, the length of the simplest descriptive code for an individual pattern determines the size of the set of all theoretically possible equally complex patterns (as in AIT, which focuses on the algorithmically relevant complexities of simplest descriptive codes) or the set of all theoretically possible equally structured patterns (as in SIT, which focuses on the perceptually relevant structural classes implied by simplest descriptive codes). The fact that descriptively simpler patterns belong to smaller structural classes (Collard & Buffart, 1983) agrees with Garner's (1962, 1970) idea of inferred subsets and his motto of "good patterns have few alternatives". For instance, the set of all imaginable squares is smaller than the set of all imaginable rectangles. In fact, in perception, the structural class to which a pattern belongs is considered to be more relevant than its precise metrical details (McKay, 1950), so that one could say that this class constitutes the generic representation of the pattern (e.g., the mental representation of a particular square primarily represents "a square" and its precise size is secondary). This suggests that a pattern should not be treated in isolation, but in reference to its structural class (Lachmann & van Leeuwen, 2005a, 2005b).

Hence, all in all, it is true that Shannon's optimal codes have a flavor of simplicity. They are shorter for more frequently occurring things, and thereby, minimize the long-term average length of nominalistic label codes over many identical and different things. However, it is crucial to distinguish this from the simplicity principle which minimizes the length of descriptive codes for individual things. Furthermore, notice that the foregoing deals with view-independent properties only. Indeed, initially, both the simplicity principle and likelihood principle focused on view-independent properties of hypothesized distal objects to predict the most likely outcome of the perceptual organization process - that is, ignoring how well hypotheses fit the proximal data. The latter issue is about view-

dependencies, and as discussed next, the inclusion of this issue boosted research on perceptual organization.

3.2. View-dependencies

Because descriptive simplicity is a fairly stable concept (see above), the assessment of complexities of hypothesized distal objects (i.e., objects as hypothesized in candidate interpretations) as such is not a big problem for the simplicity principle. For the likelihood principle, however, the assessment of their probabilities is a problem. It predicts that the most likely outcome of the perceptual organization process is the one that is also objectively most likely to be true in the world. However, despite suggestions (Brunswick, 1956), such objective probabilities in the world are unknown if not unknowable. This does not exclude that perception is guided by the likelihood principle, but it does mean that this may not be verifiable (Leeuwenberg & Boselie, 1988).

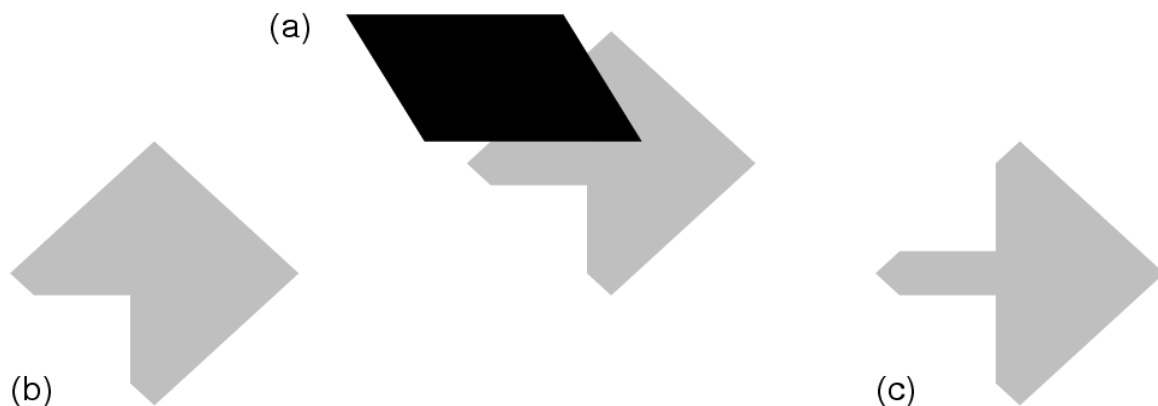


Figure 2. The pattern in (a) is readily interpreted as a parallelogram partly occluding the shape in (b) rather than the shape in (c). In this case, this preference could be claimed to occur either because, unlike the shape in (b), the shape in (c) would have to take a rather coincidental position to yield the pattern in (a), or because the shape in (b) is simpler than the shape in (c). In general, however, both factors seem to play a role.

Be that as it may, in the 1980s, proponents of the likelihood principle switched to view-dependent properties, that is, to properties that determine the degree of consistency between a candidate interpretation and the proximal stimulus (see, e.g., Gregory, 1980). For these properties, fair approximations of their objective probabilities in the world can be assessed better. This led to a debate in which advocates of one principle presented phenomena that were claimed to be explained by this principle but not by the other principle - however, advocates of the other principle were generally able to counter such arguments (see, e.g., Boselie & Leeuwenberg's, 1986, reaction to Rock, 1983, and to Pomerantz & Kubovy, 1986; Sutherland's, 1988, reaction to Leeuwenberg & Boselie,

1988; Leeuwenberg, van der Helm, & van Lier's, 1994, reaction to Biederman, 1987). The crux of this debate is illustrated by Figure 2, for which both principles - as formulated at the time - would make the correct amodal-completion prediction. That is, the simplicity principle could say that the preferred interpretation is the one in which, viewpoint independently, the completed shape is the simplest one. The likelihood principle, conversely, could say that it is the one without unlikely view-dependent coincidences of edges and junctions of the two shapes.

Both arguments seemed to be valid, and in both the simplicity paradigm and the likelihood paradigm, the result of this debate was the insight that perceptual organization requires an integrated account of both view-independent and view-dependent factors (see, e.g., Gigerenzer & Murray, 1987; Knill & Richards, 1996; Tarr & Bülthoff, 1998; van der Helm, 2000; van Lier, van der Helm, & Leeuwenberg, 1994, 1995; van Lier, 1999). For the simplicity principle, such an integration implies compliance with the MDL principle in AIT (see above), and no matter which underlying principle one adopts, it concurs with an integration of information from the ventral and dorsal streams in the brain (Ungerleider & Mishkin, 1982). These streams are believed to be dedicated to object perception and spatial perception, respectively, and an integration of view-independent and view-dependent factors can thus be said to reflect an interaction between these streams, to go from percepts of objects as such to percepts of objects arranged in space.

Hence, the past few decades showed a convergence of ideas about the factors to be included in perceptual organization. This convergence, however, does not mean that the two principles agree on how these factors are to be quantified. As explicated next in Bayesian terms, the latter issue is not just a matter of complexities versus probabilities.

3.3. Bayesian models

Thomas Bayes (1702-1761) proposed what became known as Bayes' rule (Bayes, 1763/1958). It holds that the posterior probability $p(H|D)$ of hypothesis H given data D is to be computed by multiplying the prior probability $p(H)$ of hypothesis H as such and the conditional probability $p(D|H)$ of data D given hypothesis H (it also involves a normalization factor, but this factor is currently irrelevant as it does not affect the ranking of hypotheses by their posterior probabilities).

Bayes' rule is a powerful mathematical tool to model all kinds of things in terms of probabilities (for more background information, see Feldman, Chapter 56). Its general goal is to establish a posterior probability distribution over 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. Notice, however, that Bayes' rule does not prescribe where the prior and conditional

probabilities come from (cf. Watanabe, 1969). The failure to recognize this crucial point has led to overly strong claims (see also Bowers & Davis, 2012a, 2012b). For instance, Chater (1996) claimed that the simplicity and likelihood principles in perception are equivalent, but this claim assumed implicitly - and incorrectly - that any Bayesian model automatically implies compliance with the Helmholtzian likelihood principle (van der Helm, 2000, 2011a). This may be clarified further as follows.

In Bayesian terms, the above-mentioned convergence of ideas about the factors to be included in perceptual organization means that both the likelihood paradigm and the simplicity paradigm nowadays promote an integration of priors and conditionals - where the priors refer to view-independent factors of candidate interpretations as such, while the conditionals refer to their view-dependent degree of consistency with proximal stimuli. Hence, Bayes' rule can be employed to predict the most likely outcome of the human perceptual organization process. However, for a modeler, the key question then is: where do I get the priors and conditionals from? If one wants to model perceptual organization rather than explaining it, one might subjectively choose certain probabilities, whether or not backed-up by compelling arguments (for fine examples, see Knill & Richards, 1996). This is customary in Bayesian approaches, but notice that compliance with either one of the explanatory simplicity and likelihood principles requires more specific probabilities.

The natural way to model the likelihood principle, on the one hand, is to use Bayes' rule. After all, this principle assumes that objective probabilities in the world (p_w) determine the outcome of the perceptual organization process. That is, for proximal stimulus D , the likelihood principle can be formalized in Bayesian terms by:

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

where $p_w(H)$ is the prior probability of hypothesis H , while $p_w(D|H)$ is the probability that the proximal stimulus D arises if the real distal stimulus is as hypothesized in H .

The natural way to model the simplicity principle, on the other hand, is to minimize the sum of prior and conditional complexities (just as specified for the MDL principle in AIT). However, one may also convert descriptive complexities C into artificial probabilities $p_a=2^{-C}$; these are called algorithmic probabilities in AIT (Li & Vitányi, 1997) and precisals in SIT (van der Helm, 2000). Under this conversion, minimizing the sum of prior and conditional complexities C is equivalent to maximizing the product of prior and conditional probabilities p_a . Normalization then is irrelevant, and these artificial probabilities thus imply that also the simplicity principle can be formalized in Bayesian terms, namely, by:

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

Thus, both principles can be formalized in Bayesian terms to predict the most likely outcome of the perceptual organization process. The crucial difference then still is, however, that the likelihood principle employs probabilities p_w based on the frequency of occurrence of things in the world whereas the simplicity principle employs probabilities p_a derived from the descriptive complexity of individual things.

Hence, to determine if the Bayesian formulation of the simplicity principle complies with the likelihood principle, one should assess how close the latter's objective probabilities p_w and the former's artificial probabilities p_a might be (van der Helm, 2000, 2011a). This is discussed further in the next section, but notice that a proof of equivalence of the principles is out of the question, simply because the p_w are unknown. The next two examples may illustrate various things discussed so far.

3.4. Example 1: Straight versus curved edges

The general viewpoint assumption is an assumption put forward in the likelihood paradigm (Biederman, 1987; Binford, 1981; Rock, 1983; Witkin & Tenenbaum, 1983). It holds that a proximal stimulus is interpreted assuming it does not contain features that would arise only in an accidental view of the distal stimulus. This suggests, for instance, that a proximal straight line can safely be interpreted as a distal straight edge because it can be caused by a distal curved edge only from an accidental viewpoint position. Straightness is therefore called a nonaccidental property: if such a property is present in the proximal stimulus, then it is most likely present in the distal stimulus too.

The general viewpoint assumption is indeed plausible, but notice that it derives its plausibility from favoring interpretations involving high conditional probabilities. For instance, a curved distal edge yields a straight proximal line from hardly any viewpoint, so that a straight proximal line has a low probability to occur if the curved distal edge hypothesis were true. A straight distal edge, conversely, yields a straight proximal line from nearly every viewpoint, so that a straight proximal line has a high probability to occur if the straight distal edge hypothesis were true. It is true that Pomerantz and Kubovy (1986) argued that, in the case of a straight proximal line, the preference for the straight distal edge hypothesis should be justified by showing that straight edges occur more frequently in the world than curved edges. This, however, would be a justification in terms of prior probabilities whereas, as just argued, it is justified better in terms of conditional probabilities. Yet, according to Bayes' rule, a high conditional probability may be suppressed by a low prior probability, so, it still remains to be seen if the prior probability in the world is high enough to allow for a justification within the likelihood

paradigm (Leeuwenberg et al., 1994).

3.5. Example 2: T-junctions

Each of the four configurations in Figure 3 can, in principle, be interpreted as consisting of one object or as consisting of two objects. Going from left to right, however, the two-objects interpretation (definitely preferred in a) gradually loses strength in favor of the one-object interpretation (definitely preferred in d). By way of a clever experiment involving twelve of such configurations, Feldman (2007) provided strong evidence for this. For instance, he found that, just as the configuration in a, the T-junction in b is perceived as two objects, and that, just as the configuration in d, the hook in c is perceived as one object.

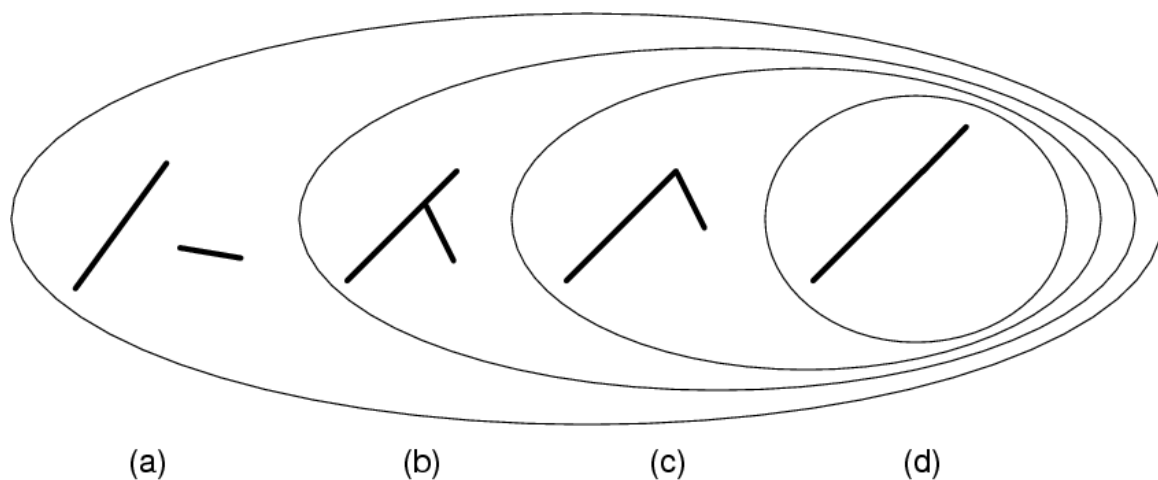


Figure 3. Four configurations that can be interpreted as consisting of one object or as consisting of two objects. Taken as one object, a simpler (i.e., more regular) one belongs to a smaller object category; taken as two objects, a simpler (i.e., less coincidental) relative position of the two objects belongs to a larger position category.

T-junctions are particularly interesting because, in many models of amodal completion, they are considered to be cues for occlusion (see, e.g., Boselie, 1994; see also van Lier & Gerbino, Chapter 11). That is, if the proximal stimulus contains a T-junction, this is taken as a strong cue that the distal scene comprises one surface partly occluded by another (see, e.g., Figure 2). However, before the visual system can infer this occlusion, it first has to segment the proximal stimulus into the visible parts of those two surfaces, and Feldman's (2007) data in fact suggest that T-junctions are cues for segmentation rather than for occlusion. That is, they trigger segmentation even when occlusion is not at hand.

To explain this, one may invoke van Lier et al.'s (1994) empirically-successful amodal-completion model. It quantifies prior complexities of

interpretations using SIT's coding model, and it quantifies conditional complexities under the same motto, namely, that complexity reflects the effort to construct things. Thus, for an interpretation, the prior complexity reflects the effort to construct the hypothesized distal objects, and the conditional complexity reflects the effort to bring these objects in the relative position given in the proximal stimulus. Notice that these conditional complexities are quantitatively equal to what Feldman (2007, 2009) called codimensions - with the difference that Feldman (who assumed uniform priors) took a high codimension to be an asset of an interpretation, whereas van Lier et al. (who assumed nonuniform priors) took a high conditional complexity to be a liability. The latter agrees with the simplicity principle, and implies the following for Figure 3.

Going from left to right, the one-object interpretation has prior complexities of 5, 4, 3, and 1 (reflecting the number of line segments and angles needed to describe each configuration as one object) and a conditional complexity of 0 in each case (i.e., no degree of positional freedom to be removed to arrive at the proximal configurations). Likewise, the two-objects interpretation has a prior complexity of 2 in each case (i.e., just two separate line segments to be described) and conditional complexities of 0, 1, 2, and 3 (reflecting the degrees of positional freedom to be removed to arrive at the proximal configurations). Hence, the one-object interpretation has posterior complexities of 5, 4, 3, and 1, respectively, and the two-objects interpretation has posterior complexities of 2, 3, 4, and 5, respectively. This explains Feldman's (2007) data that the hook is preferably interpreted as one object whereas the T-junction is preferably interpreted as two objects (see also van der Helm, 2011a).

Hence, both examples stress the relevance of an interplay between nonuniform priors and nonuniform conditionals. Notice that this still stands apart from the difference between the simplicity and likelihood principles. This difference returns in the next section.

4. The veridicality of simplicity

Evolutionary, a fair degree of veridicality in the world seems a prerequisite for any visual system to survive. The likelihood principle yields highly veridical percepts by definition, but what about the simplicity principle? It is true that Mach (1922/1959) suggested that simplicity and likelihood are different sides of the same coin; that Perkins (1976) concluded that simplest interpretations run little risk of misinterpreting stimuli; and that the MDL principle postulates that simplest interpretations are the best ones. However, it is not obvious at all that simplicity yields veridicality (see also Sober, 2002). For instance, the simplicity and likelihood principles cannot be proved to be equivalent (see

above). The next two preconsiderations set the stage for a further discussion of this issue.

4.1 Preconsideration 1: Feature extraction versus feature integration

In neuroscience, the perceptual organization process is believed to comprise three intertwined subprocesses which, together, yield integrated percepts composed of selected features (Lamme & Roelfsema, 2000; Lamme, Supèr, & Spekreijse, 1998). These subprocesses are feature extraction, feature binding, and feature selection (see next section for more details). As for feature extraction, the visual system's sensitivity to basic features such as line orientations seems to correlate with their objective probabilities of occurrence in the world (Howe & Purves, 2004, 2005; Yang & Purves, 2003, 2004). This is interesting as it suggests that the visual system's capability to extract features has adapted to the statistics in the world. This may even extend to features like symmetry, and seems to be in the spirit of the likelihood principle rather than the simplicity principle.

The simplicity principle is indeed silent about the visual system's feature extraction capability, but notice that it is in its spirit to assume that, via a two-way interaction between visual systems and the world, feature extraction mechanisms obtained sufficient evolutionary survival value (see below; see also van der Helm, Chapter 17). Currently more important, however, is that the simplicity and likelihood principles differ fundamentally regarding the selection of integrated percepts, and that the issue at stake here is not the visual system's feature extraction capability, but the veridicality of integrated percepts.

4.2. Preconsideration 2: Occamian bias in Bayesian modeling

It has been noticed that Bayesian models tend to exhibit a bias towards simplicity (MacKay, 2003), and this bias has been taken to reflect a rapprochement of the simplicity and likelihood principles (Feldman, 2009; Sober, 2002). This bias, however, has nothing to do with the Helmholtzian likelihood principle, and merely reflects a Bayesian implementation of the simplicity principle. This becomes clear if one looks closer at MacKay's (2003) explication of this bias. MacKay argued that a category of more complex instances spreads probability mass over a larger number of instances than a category of simpler instances does, so that individual instances in such a smaller category tend to get higher probabilities. This, however, presupposes (a) a correlation between complexity and category size, and (b) that every category gets an equal probability mass. These assumptions cannot be justified within the likelihood paradigm, but are in line with the simplicity paradigm.

That is, MacKay seemed to have in mind a world with objects generated, each time, by first selecting randomly a complexity category, and then by selecting randomly an instance from that category. Thus, in the first step,

every category has a same probability of being selected, and in the second step, every instance in the selected category has again a same probability of being selected. The instances in a category of complexity C can be defined by C parameters, so that 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 nothing but the simplicity paradigm's artificial probability p_a (see previous section).

4.3. The margin between simplicity and likelihood

In the just-sketched imagined world, the simplicity and likelihood principles would actually be equivalent (at least, regarding the priors). Thereby, it touches upon the heart of the veridicality issue, that is, it immediately raises the question of how close this imagined world might be to the actual world, or more general, the question of how close the two principles might be in other imaginable worlds. Because the probabilities in the actual world are unknown, the first question cannot be answered, but the second question found an answer in AIT's Fundamental Inequality (Li & Vitányi, 1997) which, in words, holds:

For any enumerable probability distribution P over things x with Kolmogorov complexities $K(x)$, the difference between the real probabilities $p(x)$ and the artificial probabilities $2^{-K(x)}$ is maximally equal to the complexity $K(P)$ of the distribution P .

An enumerable distribution is (or can, with arbitrary precision, be approximated by) a rational-valued function of two nonnegative integer arguments (examples are the uniform distribution, the normal distribution, and the Poisson distribution). Furthermore, the complexity $K(P)$ is the length of a shortest descriptive code specifying the probabilities $p(x)$, that is, it is roughly given by the number of different categories to which P assigns probabilities. In other words, the fewer different categories to be considered, the fewer different probabilities to be assigned, the simpler the probability distribution is.

The Fundamental Inequality is admittedly a very general finding. It is unknown if any actual world exhibits an enumerable distribution over things, and Kolmogorov complexity is in fact an incomputable theoretical construct. Nevertheless, this finding holds for both priors and conditionals and suggests that, depending on the probability distribution in a world at hand, the simplicity and likelihood principles might be close. The next question then is what this evidence suggests regarding the veridicality of simplest interpretations in perception.

In this respect, notice first that natural environments like jungles exhibit larger shape diversities than those exhibited by human-made environments like cities. The Fundamental Inequality then suggests that simplicity-guided visual systems yield a higher degree of veridicality in human-made environments than in natural environments. This makes

sense considering that jungle inhabitants rely on smell and sound rather than on sight. In fact, the Fundamental Inequality seems to explain why organisms tend to create environments with reduced shape diversity (Allen, 1879), that is, if visual systems indeed are guided by simplicity, then reducing shape diversity enables them to yield more veridical percepts. This would establish the above-mentioned two-way interaction between visual systems and the world (van der Helm, 2011b). To evaluate the relevance of the Fundamental Inequality in perception in more detail, one has to consider priors and conditionals separately.

First, even in human-made environments, the shape diversity may be too large to allow for a simple probability distribution. The Fundamental Inequality then suggests that the difference between prior probabilities p_w in the world and simplicity-based artificial prior probabilities p_a may well be large. In any case, there is no indication that the p_a might be veridical. Another way of looking at this is by considering structural-class sizes. That is, simpler objects (i.e., those with higher p_a) belong to smaller object categories (see Figure 3), which suggests that they probably occur with lower p_w in the world. Hence, the simplicity and likelihood principles seem far apart regarding the priors. For instance, straight edges are simpler than curved edges, but there is no reason to assume they occur more frequently.

Second, different views of a scene usually give rise to only a few qualitatively different spatial arrangements of objects. This small diversity suggests, by the Fundamental Inequality, that the difference between conditional probabilities p_w in the world and simplicity-based artificial conditional probabilities p_a may well be small, so that the p_a may well be veridical. To look at this too in another way, Figure 3 illustrates that simpler arrangements (i.e., those with higher p_a) belong to larger sets of position categories, which suggests that they probably also occur with higher p_w in the world. Hence, the simplicity and likelihood principles seem close regarding the conditionals. For instance, for the spatial arrangements in Figure 3, the conditional complexities as formally quantified by van Lier et al. (1994) are in fact basically identical to the number of coincidences one would count intuitively. Hence, taking high conditional complexities to be a liability (as the simplicity principle does) agrees with Rock's (1983) avoidance-of-coincidences principle which is in line with the general viewpoint assumption as put forward in the likelihood paradigm (see previous section).

Thus, in sum, the simplicity principle's priors are probably not veridical, but its conditionals probably are. On the one hand, this suggests that attempts to assess if the human visual system is guided by simplicity or by likelihood should focus on the priors, because the conditionals do not seem to be decisive in this respect. On the other hand, the simplicity principle's veridicality difference between priors and conditionals might

explain experiences that scenes look weird at first glance, but less so at subsequent glances. That is, by way of co-evolution, seeing organisms can usually move as well, and this allows them to get different views of a same scene to infer better what the scene entails. This inference process can be modeled neatly by a recursive application of Bayes' rule, which means that posteriors obtained for one glance are taken as priors for the next glance. This implies that the effect of the first priors fades away and that the conditionals become the decisive entities. Hence, although the simplicity principle's priors probably are not veridical, the fact that its conditionals probably are veridical seems sufficient to reliably guide actions in everyday situations. In other words, a visual system that aims at internal efficiency seems to yield, as a side-effect, an evolutionary sufficient degree of veridicality in the external world.

5. The neural realization of simplicity

The previous sections focused on the question of what is processed rather than on the question of how things are processed. That is, the simplicity and likelihood principles predict which interpretations result from the perceptual organization process, but this does not yet indicate how candidate interpretations are processed. Notice that any stimulus may give rise to a superexponential number of candidate interpretations, so that evaluating each of them separately may require more time than is available in this universe (cf. van Rooij, 2008). To allow for a tractable process, the likelihood paradigm tends to rely on heuristics (see, e.g., Hoffman, 1998), but this does not yet indicate how candidate interpretations are mentally structured and represented. The simplicity paradigm relies on descriptive coding schemes which do suggest how candidate interpretations are mentally structured and represented, but this does not yet resolve the tractability question (cf. Hatfield & Epstein, 1985).

What is clear, however, is that the simplicity principle requires a nonlinear process: in line with the law of Prägnanz, it implies that a minor change in the input may give a dramatic change in the output. This is also the case in connectionism and DST, and honoring ideas therein, findings within SIT in fact open - in an explanatory or epistemological sense (cf. Jilk, Lebiere, O'Reilly, & Anderson, 2008) - a pluralist perspective on how the brain might arrive at simplest interpretations. This is explicated in the next given context of (a) processing in the visual hierarchy in the brain, and perhaps surprising (b) quantum computing.

5.1. The visual hierarchy in the brain

As mentioned, neurally, the perceptual organization process is believed to comprise three intertwined subprocesses, namely, feature extraction, feature binding, and feature selection (see Figure 4). Together, these

subprocesses yield integrated percepts composed of selected features. For instance, the exogenous (i.e., stimulus-driven) subprocess of feature extraction - which is also called the feedforward sweep - codes more complex things in higher visual areas. Furthermore, the recurrent subprocess of feature selection selects different features from feature constellations and integrates them into percepts. Here, without excluding influences by endogenous (i.e., attention-driven) recurrent processing starting from beyond the visual hierarchy (Lamme & Roelfsema, 2000; Lamme et al., 1998; Peterson, 1994), the latter subprocess is taken to be a predominantly exogenous subprocess within the visual hierarchy (Gray, 1999; Pylyshyn, 1999). Currently more relevant, those feature constellations are thought to be the result of the exogenous subprocess of horizontal binding of similar features coded within visual areas. This subprocess seems to be mediated by transient neural assemblies which also have been implicated in the phenomenon of neuronal synchronization (Gilbert, 1992). This phenomenon is discussed next in more detail.

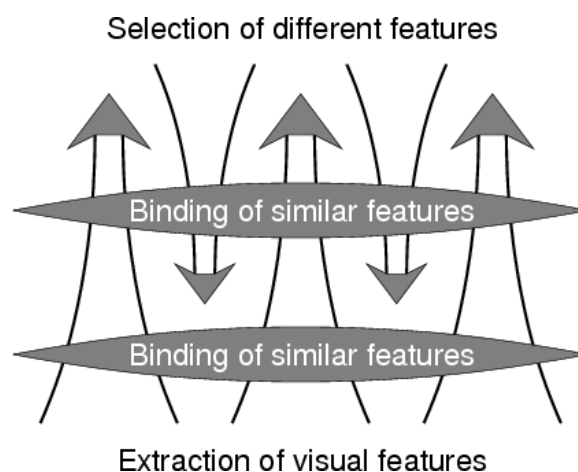


Figure 4. The process in the visual hierarchy in the brain is believed to comprise the three intertwined subprocesses of feedforward feature extraction, horizontal feature binding, and recurrent feature selection.

Neuronal synchronization is the phenomenon that neurons, in transient assemblies, temporarily synchronize their activity. Not to be confused with neuroplasticity which involves changes in connectivity, such assemblies are thought to arise when neurons shift their allegiance to different groups by altering connection strengths (Edelman, 1987), which may also imply a shift in the specificity and function of neurons (Gilbert, 1992). Both theoretically (Milner, 1974; von der Malsburg, 1981) and empirically (e.g., Eckhorn et al., 1988, 2001; Finkel, Yen, & Menschik, 1998; Fries, 2005; Gray & Singer, 1989; Salinas & Sejnowski, 2001), neuronal synchronization has been associated with cortical integration, and more general, with cognitive processing. Synchronization in the

gamma-band (30-70 Hz), in particular, has been associated with feature binding in perceptual organization.

It is true that these associations are indicative of what neuronal synchronization is involved in, but notice that they are not indicative of the nature of the underlying process. For instance, not only inside but also outside connectionism, the neural network in the brain is taken to perform parallel distributed processing (PDP). PDP, however, neither requires nor automatically implies synchronization which, therefore, is likely to subserve a form of neuro-cognitive processing that is more special than standard PDP. The question then is what this special form of processing might be.

The neural side of this question has been investigated in DST. That is, by varying system parameters, DST has yielded valuable insights into the physical conditions under which networks may exhibit synchronization (e.g., Buzsáki & Draguhn, 2004; Campbell, Wang, & Jayaprakash, 1999; Hummel & Holyoak, 2003, 2005; van Leeuwen, Steyvers, & Nooter, 1997). The point now is that SIT's simplicity approach provides complementary insights, namely, into the cognitive side of synchronization. To set the stage, the next subsection ventures briefly into the prospected application of quantum physics in computing.

5.2. Quantum computing

Classical computers work with bits. A bit represents either a one or a zero, so that a classical computer with N bits can be in only one of 2^N states at any one time. Quantum computers, conversely, are prospected to work with qubits (Feynman, 1982). A qubit can represent a one, a zero, or any quantum superposition of these two qubit states, so that a quantum computer with N qubits can be in an arbitrary superposition of up to 2^N states simultaneously. A final read-out gives one of these states, but crucially, the superposition of all these states directly affects the outcome of the read-out. Such a superposition effectively means that, until the read-out, the up to 2^N superposed states can be processed in what van der Helm (2004) called a transparallel fashion, that is, simultaneously as if only one state were concerned. Hence, compared to naive computing methods, quantum computing promises a dramatic reduction in the amount of work and time needed to complete a computing task.

Inspired by this, quantum-physical phenomena like superposition have been proposed to underlie consciousness in that they might be the source of neuronal synchronization (Penrose, 1989; Penrose & Hameroff, 2011; see also Atmanspacher, 2011). It is true that this quantum mind hypothesis does not seem tenable, because quantum-physical phenomena do not seem to last long enough to be useful for neuro-cognitive processing (Chalmers, 1995, 1997; Searle, 1997; Seife, 2000; Stenger,

1992; Tegmark, 2000). However, a cognitive form of superposition seems yet needed to account for perceptual organization (see also Townsend, Wenger, & Khodadadi, Chapter 29, and Townsend & Nozawa's, 1995, similar call for what they coined a coactive architecture yielding supercapacity). As discussed next, SIT provides such a cognitive option; it is perhaps somewhat speculative and technical, but it is also mathematically sound and neurally plausible.

5.3. The transparallel mind hypothesis

Within SIT, an algorithm has been developed to compute simplest codes of symbol strings (van der Helm, 2004). Symbol strings are not visual stimuli, but the objective of computing simplest codes raises basically the same problems. To be more specific, this algorithm relies on distributed representations of transparent holographic regularities (see van der Helm, Chapter 17), and implements the three intertwined subprocesses that are believed to take place in the visual hierarchy in the brain (see Figure 4). For instance, it implements the subprocess of feature selection by way of Dijkstra's (1959) shortest path method. This method relates SIT's algorithm to connectionist modeling because it is comparable to computer implementations, in connectionist simulations, of selection by activation spreading. A notable difference, though, is that it is not applied to one fixed network suited for all possible inputs (as in standard connectionist modeling), but to a hierarchy of input-dependent networks which represents all candidate interpretations for only the input at hand.

Such an input-dependent network on N nodes at some hierarchical level forms a superposition of up to 2^N similar regularities extracted from the previous hierarchical level. These input-dependent networks therefore find neuronal counterparts in the transient neural assemblies that are thought to be responsible for binding similar features. Moreover, such an input-dependent network is provably a hyperstring, which means that the up to 2^N superposed regularities can be hierarchically recoded in a transparallel fashion, that is, simultaneously as if only one regularity were concerned (van der Helm, 2004).

Hence, transparallel processing by hyperstrings is in fact as powerful as transparallel processing by quantum computers. A notable difference, though, is that quantum computers form a still prospected hardware-option to perform transparallel processing, whereas hyperstrings provide an already feasible software-option to perform transparallel processing on classical computers. This challenges the alleged but unproved general superiority of quantum computers over classical computers (cf. Hagar, 2011). By the way, more sophisticated computing methods usually have more application restrictions, and the vast majority of computing problems cannot benefit from either transparallel method. This does not detract from what they can do, however, and each method is bound to find its own niche.

Currently more relevant, transparallel processing by hyperstrings not only enables a tractable computation of simplest codes of symbol strings, but also provides a computational explanation of neuronal synchronization (van der Helm, 2012, 2013). That is, as said, neuronal synchronization is something else than standard PDP, and it might well be a manifestation of transparallel recoding of similar features. Whether this explanation is tenable remains to be seen, but for one thing, this pluralist picture of transient hyperstring-like neural assemblies subserving transparallel feature processing does justice to the high combinatorial capacity and speed of the human perceptual organization process.

6. Conclusions

It remains to be seen if human perceptual organization is indeed guided by the Occamian simplicity principle which aims at internal efficiency, but this chapter shows that this principle is a serious contender of the Helmholtzian likelihood principle which aims at external veridicality. The controversy between these principles is plagued by unclarities, but as reviewed, these unclarities can be resolved - enabling a clear view on their fundamental differences. One insight then is that empirical attempts to distinguish between them should focus on view-independent aspects of candidate stimulus interpretations, because view-dependent aspects do not seem to be decisive in this respect. Their functional equivalence regarding view-dependent aspects, in turn, suggests that also the simplicity principle has evolutionary survival value in that it yields sufficient veridicality in everyday situations. Furthermore, the simplicity principle's stance - that internal neuro-cognitive mechanisms tend to yield parsimonious percepts - is not only in line with Gestalt psychology but is also sustained by the computational explanation of neuronal synchronization as being a manifestation of transparallel feature processing. This explanation suggests that the simplicity principle is neurally realized by way of flexible cognitive architecture implemented in the relatively rigid neural architecture of the brain.

References

- Allen, G. (1879). The origin of the sense of symmetry. *Mind*, 4, 301-316.
- Atmanspacher, H. (2011). Quantum approaches to consciousness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <http://plato.stanford.edu>
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61, 183-193.
- Attneave, F. (1982). Prägnanz and soap-bubble systems: A theoretical exploration. In J. Beck (Ed.), *Organization and representation in perception* (pp. 11-29). Hillsdale, NJ: Erlbaum.
- Bayes, T. (1958). Studies in the history of probability and statistics: IX. Thomas Bayes' (1763) essay "Towards solving a problem in the doctrine of chances" (in modernized notation). *Biometrika*, 45, 296-315.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.
- Binford, T. (1981). Inferring surfaces from images. *Artificial Intelligence*, 17, 205-244.
- Boselie, F. (1994). Local and global factors in visual occlusion. *Perception*, 23, 517-528.
- Boselie, F., & Leeuwenberg, E. L. J. (1986). A test of the minimum principle requires a perceptual coding system. *Perception*, 15, 331-354.
- Bowers, J. S. & Davis, C. J. (2012a). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 3, 389-414.
- Bowers, J. S. & Davis, C. J. (2012b). Is that what Bayesians believe? Reply to Griffiths, Chater, Norris, and Pouget (2012). *Psychological Bulletin*, 3, 423-426.
- Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.
- Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304, 1926-1929.
- Campbell, S. R., Wang, D. L., & Jayaprakash, C. (1999). Synchrony and desynchrony in integrate-and-fire oscillators. *Neural Computation*, 11, 1595-1619.
- Chaitin, G. J. (1969). On the length of programs for computing finite binary sequences: Statistical considerations. *Journal of the Association for Computing Machinery*, 16, 145-159.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2, 200-219.
- Chalmers, D. J. (1997). *The conscious mind: In search of a fundamental theory*. Oxford, UK: Oxford University Press.
- Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, 103, 566-581.
- Collard, R. F. A., & Buffart, H. F. J. M. (1983). Minimization of structural

- information: A set-theoretical approach. *Pattern Recognition*, 16, 231-242.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1, 269-271.
- 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.
- Eckhorn, R., Bruns, A., Saam, M., Gail, A., Gabriel, A., & Brinksmeier, H. J. (2001). Flexible cortical gamma-band correlations suggest neural principles of visual processing. *Visual Cognition*, 8, 519-530.
- Edelman, G. M. (1987). *Neural darwinism: The theory of neuronal group selection*. New York: Basic Books.
- Feldman, J. (2007). Formation of visual "objects" in the early computation of spatial relations. *Perception & Psychophysics*, 69, 816-827.
- Feldman, J. (2009). Bayes and the simplicity principle in perception. *Psychological Review*, 116, 875-887.
- Feynman, R. (1982). Simulating physics with computers. *International Journal of Theoretical Physics*, 21, 467-488.
- Finkel, L. H., Yen, S-C., & Menschik, E. D. (1998). Synchronization: The computational currency of cognition. In L. Niklasson, M. Boden, & T. Ziemke (Eds.), *ICANN 98, Proceedings of the 8th International Conference on Artificial Neural Networks* (Skövde, Sweden, 2-4 September, 1998). New York, NY: Springer-Verlag.
- Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9, 474-480.
- Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York, NY: Wiley.
- Garner, W. R. (1970). Good patterns have few alternatives. *American Scientist*, 58, 34-42.
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron*, 9, 1-13.
- Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: Still alive and well. *Neuron*, 24, 31-47.
- 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.
- Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London*, B290, 181-197.
- Hagar, A. (2011). Quantum computing. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <http://plato.stanford.edu>
- Hatfield, G. C., & Epstein, W. (1985). The status of the minimum principle

- in the theoretical analysis of visual perception. *Psychological Bulletin*, 97, 155-186.
- Hochberg, J. E., & McAlister, E. (1953). A quantitative approach to figural "goodness". *Journal of Experimental Psychology*, 46, 361-364.
- Hoffman, D. D. (1998). *Visual intelligence*. New York: Norton.
- Howe, C. Q., & Purves, D. (2004). Size contrast and assimilation explained by the statistics of natural scene geometry. *Journal of Cognitive Neuroscience*, 16, 90-102.
- Howe, C. Q., & Purves, D. (2005). Natural-scene geometry predicts the perception of angles and line orientation. *Proceedings of the National Academy of Sciences USA*, 102, 1228-1233.
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220-264.
- Hummel, J. E., & Holyoak, K. J. (2005). Relational reasoning in a neurally-plausible cognitive architecture: An overview of the LISA project. *Current Directions in Cognitive Science*, 14, 153-157.
- Jilk, D. J., Lebiere, C., O'Reilly, C., & Anderson, J. R. (2008). SAL: an explicitly pluralistic cognitive architecture. *Journal of Experimental & Theoretical Artificial Intelligence*, 20, 197-218.
- Knill, D. C., & Richards, W. (Eds.) (1996). *Perception as Bayesian inference*. Cambridge, MA: Cambridge University Press.
- 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.
- Lachmann, T., & van Leeuwen, C. (2005a). Individual pattern representations are context-independent, but their collective representation is context-dependent. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 58, 1265-1294.
- Lachmann, T., & van Leeuwen, C. (2005b). Task-invariant aspects of goodness in perceptual representation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 58, 1295-1310.
- 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.
- 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.
- 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. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, 76, 216-220.
- Leeuwenberg, E. L. J. (1971). A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, 84, 307-349.
- Leeuwenberg, E. L. J., & Boselie, F. (1988). Against the likelihood principle in visual form perception. *Psychological Review*, 95, 485-491.
- 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.
- Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications* (2nd ed.). New York: Springer-Verlag.
- Mach, E. (1959). *The analysis of sensations and the relation of the physical to the psychical*. New York: Dover. (Original work published 1922)
- 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)
- Milner, P. (1974). A model for visual shape recognition. *Psychological Review*, 81, 521-535.
- Penrose, R. (1989). *The emperor's new mind: Concerning computers, minds and the laws of physics*. Oxford, UK: Oxford University Press.
- Penrose, R., & Hameroff, S. (2011). Consciousness in the universe: Neuroscience, quantum space-time geometry and Orch OR theory. *Journal of Cosmology*, 14.
- Perkins, D. (1976). How good a bet is good form? *Perception*, 5, 393-406.
- Peterson, M. A. (1994). Shape recognition can and does occur before figure-ground organization. *Current Directions in Psychological Science*, 3, 105-111.
- Pomerantz, J., & Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 36-1-36-46). New York: Wiley.
- Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case of impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341-423.
- Rissanen, J. J. (1978). Modelling by the shortest data description. *Automatica*, 14, 465-471.
- Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
- Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience*, 2, 539-550.
- Searle, J. R. (1997). *The mystery of consciousness*. New York: The New

York Review of Books.

- Seife, C. (2000). Cold numbers unmake the quantum mind. *Science*, 287, 791.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423, 623-656.
- Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 79, 369-382.
- Sober, E. (2002). What is the problem of simplicity? In H. Keuzenkamp, M. McAleer, & A. Zellner (Eds.), *Simplicity, inference, and econometric modelling* (pp. 13-32). Cambridge, UK: Cambridge University Press.
- Solomonoff, R. J. (1964a). A formal theory of inductive inference, Part 1. *Information and Control*, 7, 1-22.
- Solomonoff, R. J. (1964b). A formal theory of inductive inference, Part 2. *Information and Control*, 7, 224-254.
- Stenger, V. (1992). The myth of quantum consciousness. *The Humanist*, 53, 13-15.
- Sutherland, S. (1988). Simplicity is not enough. In B. A. G. Elsendoorn & H. Bouma (Eds.), *Working models of human perception* (pp. 381-390). London: Academic Press.
- Tarr, M. J., & Bühlhoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67, 1-20.
- Tegmark, M. (2000). Importance of quantum decoherence in brain processes. *Physical Review E*, 61, 4194-4206.
- Townsend, J.T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, 39, 321-359.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge, MA: MIT Press.
- van der Helm, P. A. (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, 126, 770-800.
- van der Helm, P. A. (2004). Transparallel processing by hyperstrings. *Proceedings of the National Academy of Sciences USA*, 101 (30), 10862-10867.
- van der Helm, P. A. (2011a). Bayesian confusions surrounding simplicity and likelihood in perceptual organization. *Acta Psychologica*, 138, 337-346.
- van der Helm, P. A. (2011b). The influence of perception on the distribution of multiple symmetries in nature and art. *Symmetry*, 3, 54-71.
- van der Helm, P. A. (2012). Cognitive architecture of perceptual organization: From neurons to gnosons. *Cognitive Processing*, 13, 13-40.
- van der Helm, P. A. (2013). *Simplicity in vision: A multidisciplinary account of perceptual organization*. Cambridge, UK: Cambridge University Press.

- van Leeuwen, C., Steyvers, M., & Nooter, M. (1997). Stability and intermittency in large-scale coupled oscillator models for perceptual segmentation. *Journal of Mathematical Psychology, 41*, 319-344.
- van Lier, R. (1999). Investigating global effects in visual occlusion: From a partly occluded square to a tree-trunk's rear. *Acta Psychologica, 102*, 203-220.
- 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.
- van Lier, R. J., van der Helm, P. A., & Leeuwenberg, E. L. J. (1995). Competing global and local completions in visual occlusion. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 571-583.
- van Rooij, I. (2008). The tractable cognition thesis. *Cognitive Science, 32*, 939-984.
- von der Malsburg, C. (1981). *The correlation theory of brain function*. Internal Report 81-2, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany.
- von Helmholtz, H. L. F. (1962). *Treatise on physiological optics* (J. P. C. Southall, Trans.). New York: Dover. (Original work published 1909)
- Watanabe, S. (1969). *Knowing and guessing*. New York: Wiley.
- Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie, 12*, 161-265.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt [On Gestalt theory]. *Psychologische Forschung, 4*, 301-350.
- Witkin, A. P., & Tenenbaum, J. M. (1983). On the role of structure in vision. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 481-543). New York: Academic Press.
- Yang, Z. Y., & Purves, D. (2003). Image/source statistics of surfaces in natural scenes. *Network-computation in neural systems, 14*, 371-390.
- Yang, Z. Y., & Purves, D. (2004). The statistical structure of natural light patterns determines perceived light intensity. *Proceedings of the National Academy of Sciences USA, 101*, 8745-8750.