The resurrection of simplicity in vision

Peter A. van der Helm

Nijmegen Institute for Cognition and Information

University of Nijmegen

Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

peterh@nici.kun.nl

Wordcount (text without references): ± 2060 **words**

The resurrection of simplicity in vision

Peter A. van der Helm

University of Nijmegen

Since early in the 20th century, simplicity has been considered a relevant factor in visual form and object perception, albeit with some ups and downs. Since the 1990s, other sciences have also shown an interest in simplicity as a driving modelling factor. This recent interest has been triggered by intriguing findings in a mathematical research line, called algorithmic information theory (AIT), that started in the mid 1960s. As I argue here, these AIT findings support but cannot replace the independent perceptual research line that started in the early 1950s with Hochberg and McAlister's (1953) information-theoretic simplicity idea.

In the aftermath of Shannon's (1948) ground-breaking work in classical information theory, Hochberg and McAlister (1953) proposed modelling the visual system as selecting the most simple interpretation of a proximal stimulus. More precisely, they proposed:

> 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)

This proposal is nowadays known as the *simplicity principle*, and it marks the beginning of a paradigm shift from probabilistic to descriptive information.

Probabilistic information is the central concept in classical information theory. After Nyquist (1924) and Hartley (1928), Shannon (1948) employed the following probabilistic quantification of the information load of an object:

if an object x occurs with probability $p(x)$, then its information load is $-\log_2 p(x)$.

In communication science, the term "object" would refer to a transmittable message and, in vision science, it would refer to a possible interpretation of a visual stimulus, that is, a message that the visual system could transmit to higher cognitive levels.

Hence, probabilistic information starts from known probabilities in terms of frequencies of occurrence: The more often an object is known to occur, the smaller its information load is. Tribus (1961) called the quantity $-\log_2 p(x)$ the *surprisal*: The more often an object is known to occur, the lower the surprise is when it does occur. In vision science, the idea of starting from known probabilities was put forward by von Helmholtz (1909/1962). His *likelihood principle* proposes modelling the visual system as selecting the interpretation with the highest probability of specifying correctly the actual distal stimulus (cf. Hochberg, 1968, p. 89). This sounds attractive because it suggests that vision is highly veridical.

However, an individual object can be assigned a probability only by virtue of knowledge about the size of the class to which this object is taken to belong, that is, only by virtue of knowledge about the frequencies of occurrence of instantiations of this class. The likelihood principle is silent about what the classes are, but even if the classes would be known, how could vision or vision science have access to the real frequencies of occurrence in this world? Ecological surveys (cf. Brunswick, 1956) might yield frequencies of occurrence, but these are frequencies of occurrence of percepts and precisely these percepts need explaining. That is, if one takes the latter frequencies of occurrence not just to fit data but also to explain data, then one falls into the circular explanation that "we see what we see" (cf. Hoffman, 1996).

Descriptive information is a concept that circumvents the foregoing problems by starting from the regularity in the internal structure of individual objects. As Hochberg and McAlister (1953) put it, the descriptive information load, or *complexity* $c(x)$, of an individual object x is:

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

Hence, in total, the simplicity principle proposes modelling the visual system as selecting the interpretation with the shortest descriptive code, that is, with the shortest reconstruction recipe for the stimulus. In general, shortest codes are codes that capture a maximum of regularity in objects.

The relevance of simplicity in perceptual organization had already been demonstrated by the early 20th century Gestaltists. In fact, the simplicity principle can be seen as an information-theoretic version of Koffka's (1935) *Law of Prägnanz* which associated simplicity with energetically-stable internal states of the visual system. In contrast to the early Gestaltists, who appealed to an intuitive concept of simplicity, Hochberg and McAlister's proposal paved the way for formal specifications of simplicity.

This rethinking, in the 1950s, of the role of simplicity in perceptual organization (see also Attneave, 1954) triggered, in the 1960s, the development of various coding models (e.g., Leeuwenberg, 1968; Restle, 1970; Simon & Kotovsky, 1963; Vitz & Todd, 1969). These coding models were fairly successful empirically but also created a new problem: Every coding model proposed another coding language yielding language-dependent complexities, so, which model is the right one? After having compared half a dozen of these coding models, Simon (1972) formulated this new problem as follows:

> if an index of complexity is to have significance for psychology, then the encoding scheme itself must have some kind of psychological basis. (p. 371)

In the years that followed, no such psychological basis was proposed and, as a consequence, vision science lost much of its interest in the simplicity principle.

This loss of interest was understandable but, in a sense, also curious. Simon's (1972) conclusion was based on his empirical finding that simplicity appeared to be a stable concept: Complexities $c(x)$ may vary in absolute value with the employed coding language, but are yet highly correlated across coding languages. In the mid 1960s, in mathematics, a theoretical finding with the same thrust, namely, the *Invariance Theorem* (Kolmogorov, 1965; Solomonoff, 1964), did not decrease but increased the interest in simplicity and lead to the rise of algorithmic information theory (AIT; see Li & Vitányi, 1997).

A central topic in AIT is the relation between probabilistic and descriptive information. A major finding in AIT is that a good alternative for the often unknown real probability $p(x)$ of object x is provably provided by starting from the complexity $c(x)$ of x and by taking $2^{-c(x)}$ as an artifical probability of x. (See also Leeuwenberg & Boselie (1988) who, in vision science, proposed a similar derivation of probabilities from complexities.)

Van der Helm (2000) called the quantity $2^{-c(x)}$ the *precisal*: It is higher if $c(x)$ is lower, and a lower $c(x)$ implies a more precise specification of x, that is, a specification that classifies x as belonging to a smaller subset of all possible objects. (Garner's (1970) motto that "good patterns have few alternatives" has the same thrust.) To be clear, a precisal does not refer to a real frequency of occurrence in this world. It refers to a frequency of occurrence in a world filled by an object generator that, first, randomly selects a class of objects with the same complexity and, then, randomly produces an object from this class. Because class size correlates with complexity, a particular simple object thus has a higher probability of being produced than a particular complex object.

The AIT finding that precisals form a good alternative for real probabilities lead, in the 1990s, to a growing interest in simplicity in various research domains. As for the relevance of AIT to vision, however, some reserve is in order. First, AIT does not provide a concrete coding language, let alone a psychological basis as called for by Simon (1972). Second, AIT does not take account of MacKay's (1950) perceptually relevant distinction between metrical and structural information. This implies, for instance, that squares and rectangles may get the same or different complexities, depending on the numerical values of the edge lengths. Third, in AIT, the precisal of an object reflects the size of the class of all different objects with the same complexity, but this form of object classification does not seem to have perceptual relevance.

Hence, without a psychologically compelling account of these issues, AIT findings cannot be translated meaningfully to vision. Fortunately, in the

1970s and 1980s, a few vision scientists kept the simplicity flame burning, most prominently within structural information theory (SIT). SIT arose from Leeuwenberg's (1968) coding model that employs complexities reflecting structural information only (see van der Helm, van Lier, & Leeuwenberg, 1992). In reponse to Simon (1972), this coding model nowadays has an adequate psychological basis which comprises a novel formalization of visual regularity that (a) underlies the choice of the complexity metric and the regularity-capturing coding rules, and (b) proved itself by way of an empirically successful model of regularity detection (van der Helm & Leeuwenberg, 1991, 1996, 1999, 2004).

Furthermore, in line with Garner's (1962) seminal idea of inferred subsets, SIT takes an individual object to belong to a *structural class*, that is, the class of all metrically different objects with not only the same structural complexity but also the same structure (Collard & Buffart, 1983). Therefore, in SIT, an object's precisal reflects the size of the object's structural class (van der Helm, 2000). Only given such a psychologically motivated derivation of precisals, the earlier-mentioned AIT finding can be translated meaningfully to vision. That is, this AIT finding then suggests that vision, if guided by precisals, is fairly veridical in many worlds (i.e., rather than highly veridical in only one world).

The foregoing shows that AIT supports but cannot replace the perceptual research line that Hochberg and McAlister (1953) started. In fact, this line still moves on, for instance, within the following Bayesian framework.

First, by going from complexities $c(x)$ to precisals $2^{-c(x)}$, selecting the most simple interpretation H for a proximal stimulus D transforms into selecting the interpretation H that, in Bayesian style, maximizes the posterior precisal $p(H|D) = p(H) * p(D|H)$. Here, the prior precisal $p(H)$ is derived from the (viewpoint-independent) complexity of the distal stimulus as hypothesized in H, and the conditional precisal $p(D|H)$ is derived from the (viewpoint-dependent) complexity of the proximal stimulus D if H would be true.

Second, based on van Lier, van der Helm, and Leeuwenberg's (1994) empirically successful distinction between viewpoint-independent and viewpoint-dependent complexities in amodal completion, van der Helm (2000) argued that conditional precisals are close to real conditional probabilities. This is relevant because the conditionals are decisive in the everyday perception of a moving observer who recursively applies Bayes' rule to interpret a growing sample of proximal stimuli of the same distal scene. After each recursion step, the just obtained posterior is taken as the prior in the next recursion step, which implies that the effect of the first priors fades away and that the conditionals become the decisive entities. Hence, if conditional precisals are indeed close to real conditional probabilities, then one may just as well use precisals instead of real probabilities, especially if the latter are unknown. Remarkably, various current Bayesian models of perception (see, e.g., those in Knill & Richards, 1996) claim to use real probabilities but actually use precisals.

Third, recent correspondence with Julian Hochberg triggered a Bayesian picture of the link between an autonomous, simplicity-guided, visual system and other cognitive faculties. To answer Hochberg's (1982) question in general terms, this picture starts with a proximal stimulus with a size that, temporally, spans about 30 ms (Leeuwenberg, Mens, & Calis, 1985) and that, spatially, is determined by acuity and conspicuity (e.g., extra-foveally, a higher conspicuity may compensate for the lower acuity). Then, in Bayesian style, hypotheses about the distal stimulus are ranked on the basis of their prior precisals which, via multiplication with their conditional precisals, are weighed by the degree of consistency between the hypothesized distal stimuli and the proximal stimulus. The prior ranking reflects a form of object perception, as it is determined by object-centered representations of hypothesized distal objects. The conditional weighing reflects a form of space perception, as it is determined by relative positions of the viewer and hypothesized distal objects.

This distinction between priors and conditionals agrees with the functional distinction between the ventral (”what”) and dorsal (”where”) pathways in the brain (Ungerleider & Mishkin, 1982). The idea now is that the Bayesian integration of these perceptual priors and conditionals guides the hypotheses to higher cognitive levels where their ranking may be weighed by further conditional factors, such as the viewer's knowledge and intentions related to a task to be performed, yielding a final ranking in the situation at hand.

Hence, to conclude, Hochberg and McAlister's (1953) simplicity idea has lead to a conception of vision as a fairly reliable, autonomous, source of knowledge about the external world. In this conception, vision performs an unconscious inference on the basis of knowledge-free precisals, the output of which is enriched by a gradually more and more conscious inference on the basis of internally available contextual information.

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*, 183–193.

Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Collard, R. F. A., & Buffart, H. F. J. M. (1983). Minimization of structural information: A set-theoretical approach. *Pattern Recognition, 16*, 231–242.

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.

Hartley, R. V. L. (1928). Transmission of information. *Bell Systems Technology Journal, 7*, 535–563.

Hochberg, J. E. (1968). *Perception.* Englewood Cliffs, NJ: Prentice-Hall.

Hochberg, J. E. (1982). How big is a stimulus? In J. Beck (Ed.), *Organization and representation in perception* (pp. 191-217). Hillsdale, NJ: Erlbaum.

Hochberg, J. E., & McAlister, E. (1953). A quantitative approach to figural "goodness". *Journal of Experimental Psychology, 46*, 361–364.

Hoffman, D. D. (1996). What do we mean by "The structure of the world"? In D. K. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 219–221). Cambridge, MA: Cambridge University Press.

Knill, D. K., & Richards, W. (Eds.) (1996). *Perception as Bayesian inference.* Cambridge, MA: Cambridge University Press.

Koffka, K. (1935). *Principles of gestalt psychology.* London: Routledge & Kegan Paul.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems in Information Transmission, 1*, 1–7.

Leeuwenberg, E. L. J. (1968). *Structural information of visual patterns: an efficient coding system in perception.* The Hague: Mouton.

Leeuwenberg, E. L. J., & Boselie, F. (1988). Against the likelihood principle in visual form perception. *Psychological Review, 95*, 485–491.

Leeuwenberg, E. L. J., Mens, L., & Calis, G. (1985). Knowledge within perception: Masking caused by incompatible interpretation. *Acta Psychologica, 59*, 91–102.

Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications* (2nd ed.). New York: Springer-Verlag.

MacKay, D. (1950). Quantal aspects of scientific information. *Philosophical Magazine, 41*, 289–301.

Nyquist, H. (1924). Certain factors affecting telegraph speed. *Bell Systems Technology Journal, 3*, 324–346.

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review, 77*, 481–495.

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.

Simon, H. A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review, 70*, 534–546.

Solomonoff, R. J. (1964). A formal theory of inductive inference. *Information and Control, 7*, 1–22, 224–254.

Tribus, M. (1961). *Thermostatics and Thermodynamics.* Princeton, NJ: Van Nostrand.

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., & Leeuwenberg, E. L. J. (1991). Accessibility, a criterion for regularity and hierarchy in visual pattern codes. *Journal of Mathematical Psychology, 35*, 151–213.

van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review, 103*, 429–456.

van der Helm, P. A., & Leeuwenberg, E. L. J. (1999). A better approach to goodness: Reply to Wagemans (1999). *Psychological Review, 106*, 622–630.

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.

van der Helm, P. A., van Lier, R. J., & Leeuwenberg, E. L. J. (1992). Serial pattern complexity: Irregularity and hierarchy. *Perception, 21*, 517–544.

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.

Vitz, P. C., & Todd, R. C. (1969). A coded element model of the perceptual processing of sequential stimuli. *Psychological Review, 76*, 433–449.

von Helmholtz, H. L. F. (1962). *Treatise on Physiological Optics* (J. P. C. Southall, Trans.) New York: Dover. (Original work published 1909)