Perception as Bayesian Inference
DC Knill, WA Richards, eds
New York: Cambridge University Press, 1996

Ideal observers, real observers, and the return of Elvis

Ronald A. Rensink

(Commentary on Chapter 6: "Implications of a Bayesian formulation of visual information for processing of psychophysics", DC Knill, D Kersten, P Mamassian)

Knill, Kersten, & Mamassian (Chapter 6) provide an interesting discussion of how the Bayesian formulation can be used to help investigate human vision. In their view, computational theories can be based on an ideal observer that uses Bayesian inference to make optimal use of available information. Four factors are important here: the image information used, the output structures estimated, the priors assumed (i.e., knowledge about the structure of the world), and the likelihood function used (i.e., knowledge about the projection of the world onto the sensors). Knill & Kersten argue that such a framework not only helps *analyze* a perceptual task, but can also help investigators to *define* it. Two examples are provided (the interpretation of surface contour and the perception of moving shadows) to show how this approach can be used in practice.

As the authors admit, most (if not all) perceptual processes are ill-suited to a "strong" Bayesian approach based on a single consistent model of the world. Instead, they argue for a "weak" variant that assumes Bayesian inference to be carried out in modules of more limited scope. But how weak is "weak"? Are such approaches suitable for only a few relatively low-level tasks, or can they be applied more generally? Could a weak Bayesian approach, for example, explain how we would recognize the return of Elvis Presley?

The formal modelling of human perception

To help get a fix on things, it is useful to examine the fate of an earlier attempt to formalize human perception: the application of information theory. It was once hoped that this theory—a close cousin of the Bayesian formulation—would provide a way to uncover information-handling laws that were largely independent of physical implementation. In this approach, the human nervous system was assumed to have communication channels of limited bandwidth; if data transfer were sufficiently slow, delays in reaction times would be measurable. By mapping out delays as a function of the probabilities of alternative inputs (i.e., the information they contained), it was hoped that the channel bandwidths could be determined, along with the amount of information associated with each stimulus (see, e.g., Gregory, 1986; Pierce, 1980).

As it turned out, information theory did not live up to these hopes. Although in principle it provided a general framework in which to explore information handling, in practice it was relevant only for highly-constrained tasks where relatively little practice was allowed, where stimuli were simple, and where strategy could not be used (see, e.g., Gregory, 1986).

What went wrong? In retrospect, it is clear that at least three factors were at play:

(i) The basic units in the perceiver were ill-defined. Not only were these units entirely internal, but they were unstable: new, more efficient units could be learned with practice (i.e., "chunking").

(ii) The probabilities attached to these units were ill-defined. If something occurred only once, what probability would it be assigned? How would probabilities be updated when new input was received? Frequencies might be used, but under what condition were the counts "reset"?

(iii) The structures actually used could not be controlled. Although the members of the input set might be clearly defined, attentional effects could restrict the units being used or even bring in extraneous ones (Gregory, 1986).

Looked at more generally, it is evident that formalisms such as those based on information theory (or on Bayesian inference) implicitly constrain a process, since its structure must conform to the structure of the formalism if relevance is to be maintained (see, e.g., Pierce, 1980). In particular, at least three conditions must be met:

(i) The basic terms can be reliably mapped to structures in the perceiver.

(ii) The relations between the terms (including probabilistic ones) match the relations between the corresponding structures in the perceiver.

(iii) The resulting structures are independent of all other factors (i.e., independent of context).

From the ideal to the real

Knill *et al* manage to avoid many of the difficulties that plagued information-theoretic approaches, since unstable subjective quantities are now replaced with stable objective ones. But this alone does not guarantee the suitability of their ideal observer, for a real observer is constrained by much more than just available information. First of all, *physical* limitations on available time and space constrain the amount of information that can be stored and transformed. Next, *biological* factors (both developmental and phylogenetic) constrain process formation--it might not be possible to have priors and likelihoods wired into the neural hardware, much less learn them from experience. Finally, *ecological* factors limit the kinds of tasks that are carried out. It is only important that an organism do the right thing; this does not necessarily require Bayesian inference.

Given these considerations, what kinds of perceptual process might be suited to an ideal-observer analysis? To begin with, "ideal" inputs and outputs must correspond to structures in the real observer. But such correspondences can be difficult to establish. For example, the external world might serve as its own representation, being sampled whenever detailed information is required (O'Regan, 1992). If the return of Elvis were recognized in this way (e.g., by verifying a few key image properties), it would be difficult to determine exactly what was involved. Furthermore, the combinatorial explosion of possible objects and events (see e.g., Tsotsos, 1987) means that—for the most part—increasingly complex structures are increasingly rare. If input and output structures are formed via learning, large differences could then arise between observers with different histories. Space limitations might also force rarely-used structures to be discarded, leading to history-dependent differences within observers. The existence of stable, well-defined structures common to all observers is therefore not guaranteed, especially at higher processing levels.

It can also be difficult to determine the kind of inference actually used. Optimal information usage does not necessarily imply Bayesian inference: other techniques exist (such as those based on signal-detection theory), and these cannot be easily (if at all) reconciled with Bayesian inference (Gigerenzer et al., 1989). Thus, for example, it would be necessary to show that recognizing the return of Elvis does not simply require Elvis-detection of some kind, but requires determining the posterior probability of this event. And because Bayesian priors are based on event frequencies, it would also be necessary to explain how priors would be assigned to a once-only event never observed before. Furthermore, there is also the matter of computational resources. The lowest levels of visual processing involve the detection of photons, a relatively simple task best analyzed using signal-detection theory (see, e.g., Barlow, 1962). Indeed, detection tasks are common in low-level vision, and it may be that the time and space limitations typical of these levels prohibit the use of (more sophisticated) Bayesian inference.

A final consideration concerns the priors. Given some well-defined prior probability of encountering Elvis, this value would certainly be higher in Tennessee than in Finland. More generally, priors are often functions of several variables--indeed, Kersten & Knill use such a "conditional" prior in their analysis of contour interpretation. Unless the prior is a true invariant, information will be thrown away by use of a context-free prior (essentially an average over all conditions). But using a context-dependent prior requires knowing the context. If this involves other inferred properties, further inference will be required, which in turn might require still more inference, and so on. This could lead to a set of strongly-coupled equations in which nothing can be determined until everything has been determined. The application of ideal observers must therefore be limited to tasks involving priors that are invariant, or at least that depend upon each other in a non-recursive way.

Problems and possibilities

Insofar as a task avoids ill-defined structures and resource-limited processes, and has well-defined priors that are relatively invariant (at least under some set of conditions), there is a chance that it can be analyzed by a weak Bayesian approach. The tasks discussed by Knill & Kersten are exactly of this type. Subjects in their experiments are given ample time to respond, and the structures involved are relatively simple ones (transparent or painted surfaces, shadows) that are meaningful and that the subjects have often seen before.

But notice that complications arise even here. For example, even the task of interpreting surface contours is too complex to be handled in its most general form. Instead, analysis is restricted to a few "modes" corresponding to special contexts that are particularly easy to handle. Knill & Kersten make the important point that their approach allows generality to be gradually increased, but it is not clear how much generalization can be done in practice. The authors also show how knowledge about the priors can be gradually refined by use of "well-formed" constraints, but again, it is not clear how far this can be taken. Their approach therefore runs the risk not only of being unsuitable for processes at the upper and lower ends of the processing stream, but also of being unsuitable for many interesting processes in between.

However, all is not lost. Although time and space limitations may intrude upon some processes, these are not always arbitrary "tricks", but rather can be graceful adaptations that make good use of available resources (see e.g., Enns & Rensink, 1991). Many take the form of rapid processes specialized for particular modes of the priors. Thus, although it might not always be possible to generalize computational models in the way envisioned by Knill & Kersten, it might still be possible to make use of their framework. In particular, a Bayesian analysis of a world might uncover modes corresponding to common contexts (e.g., direction of lighting from above) which could then be used as bases for the specialized processes. It is important to notice the shift of emphasis here--specialized problems (and processes) would no longer be convenient methodological stages, but would be true entities in their own right. In other words, ontology would recapitulate methodology.

Similar considerations apply to the variations in individual observers. These variations could be captured in a natural way by relatively coarse devices, such as the monotonicity relationship used by Knill & Kersten in their model of contour interpretation, or the lattice framework of Jepson & Richards (1992). Again, this coarseness would not just be a methodological convenience, but would reflect the true variability of the process (due to accidents of history, variations in values, etc.). Note that this variability could be more than simple variation in the values of priors and likelihoods--it could also include variation in the structures used, or the procedures applied. Taking this view, of course, does not rule out a progressive refinement of priors. But such refinement will often stop well short of a quantitative formulation. A systematic comparison of the variations present at different levels might then serve as a new source of insight into the relation between the observer and its environment.

References

Barlow, H.B. (1962). A method of determining the overall quantum efficiency of visual discriminations. *Journal of Physiology*, **160**, 155-168.

Enns, J.T., & Rensink, R.A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, **98**, 335-351.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., and Krüger, L. (1989). *The Empire of Chance.* Cambridge: Cambridge University Press.

Gregory, R.L. (1986). "Whatever happened to information theory?" In *Odd Perceptions*. London: Routledge.

Jepson, A., and Richards, W. (1992). A Lattice Framework for Integrating Vision Modules. *IEEE Transactions on Systems, Man, and Cybernetics*, **22**, 1087-1096.

O'Regan, J.K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, **46**, 461-488.

Pierce, J.R. (1980). *An Introduction to Information Theory: Symbols, Signals, and Noise*, 2nd edition. New York: Dover.

Tsotsos, J.K. (1987). A "Complexity Level" Analysis of Vision. *International Journal of Computer Vision*, **1**, 346-355.