A Biased View of Perceivers

Commentary on 'Observer theory, Bayes theory, and psychophysics,' by B. Bennett, et al.

Allan D. Jepson
University of Toronto

Jacob Feldman
Rutgers University

March 14, 1995

1 Observers and Theorists

The target chapter[1] introduces the notion of an observer as an entity which applies a given Bayesian theory to the task of inferring world properties from sense data. In fact, most of this volume is concerned with observers in one form or another. The basic idea is that the perceptual system uses a model of the current world context, that is, a 'domain theory,' to reduce the ambiguity inherent in the sense data and to select particular interpretations for the state of the world. Example domain theories are: the qualitative probabilistic models used for the motion of a ball in a box in Chapter 4; the preference based scheme for the interpretation of handle shape and orientation discussed in Chapter 6; the Bayesian models for shape from texture and shading used in Chapter 10; and the cost function formulation proposed for the interpretation of shape, illumination, and pigmentation in Chapter 11. In order to include such alternative forms of world models we use the term 'observer' to refer loosely to any entity which applies a *fixed* domain theory to the task of inferring world properties from sensory data.

Our bias is to view the perceptual system as a theorist in its own right, not just an observer. We claim that in order to understand perception we need to understand the way this theorist works, that is: how it selects a theory for a new domain, how it applies a theory to choose particular interpretations for sense data, how it compares different theories, and how it revises and learns theories under the influence of new information. In contrast an observer is concerned with only one of these aspects, namely how a fixed theory is applied.

The distinction between theorist and observer is perhaps blurred by notions such as Yuille's competitive priors (see Chapter 8 or [8]). Each prior may lead to a single distinct (Bayesian) observer, and the selection of one prior in favor of others can be thought of as a theory selection process. Thus the overall system can be thought of as a theorist. On

[1]This commentary is to appear in the book "Perception as Bayesian Inference" edited by D. Knill and W. Richards. See Bennett et al, J. Math. Psych. 37(2), 1993, pp. 220-240, for material related to the target chapter.

the other hand, the same sort of process can be captured in a composite model for which the selection of the appropriate prior for a new piece of data is dictated by a new model parameter, a so called 'hyper- parameter.' Thus the system can be also be thought of as an observer. Which one is it? We choose to resolve this issue on pragmatic grounds, deciding in favor of a theorist so long as there is interesting structure within the world model that can be understood in terms of theory selection, revision, comparison, and/or learning; these are just the characteristic tasks of a theorist beyond those of an observer. We would thus consider a system using Yuille's competitive priors, with a well specified means of selecting which prior to use, as a theorist, though it may be a particularly simple form of theorist with only a small set of built-in theories to choose from. Our belief is that the perceptual system is a much more sophisticated theorist than this (see Section 3 below), and therefore the appropriate categorization of such borderline cases is not critical to our argument.

We note in passing that Observer Theory, as it is formulated in the target chapter, has a role to play within many theorists. In particular, if a theorist is to consider Bayesian domain theories then, in order to apply a particular domain theory, the theorist must be able to derive (at least some of) the consequences of the candidate theory. A specific example is the ball-in-a-box model discussed in Chapter 4. There one needs to evaluate the posterior probability of various interpretations, given a prior in the form of a (qualitative) mixture model, complete with Dirac measures. This ability to derive properties of the posterior distribution when the prior probability is a general measure, and the observations are measurable sets, is precisely the domain of Observer Theory. As such, we believe that Observer Theory represents an important part of any rigorous mathematical theory of perception. However, as we explain below, our central point in this commentary is that Observer Theory does not encompass the additional structure of a theorist, and thus it is not sufficient on its own to account for perception.

An indication of the sort of phenomena we are attempting to explain, and which go beyond the scope of Observer Theory, is given by Feldman's experiment [3]. This experiment indicates that humans are capable of generating a novel domain theory from just a single example. In order for a theorist to be able to do this, it must have strong a priori constraints on the forms of theories it will entertain. That is, the system must be sufficiently biased so that, given only a handful of examples, a preferred domain theory can be selected from the set of all possible theories that the system can express. The critical element here is this bias, which can involve restricting the set of concepts the theorist can use, restricting the language for expressing theories, and/or applying a priori preferences for some theories over others [5]. It is primarily these components which determine how a theorist will revise old domain theories and learn new ones. This is illustrated with a concrete example in Section 3 below, but first, as motivation, we consider an alternative hypothesis for the structure of our perceptual systems which is along the lines of Observer Theory.

2 The homunculus observer

In stark contrast to our proposal for viewing perception as a process of both theory formation and application, one might instead attempt to consider the whole of the perceptual system as just a single observer (as defined in the target chapter). In particular, the transduced stimulus could be taken as the input premises, with the posterior probability distribution for the scene provided over the conclusion space. The capacity for storing and processing information within an Observer Theory framework would not be strained, nor would it be strained if we considered instead the entire past history of stimuli as a possible input space. The basic structural restrictions of an observer, namely that the prior be a positive measure and that various mappings are measurable, easily encompass such a formulation. In other words, Bennett and his coauthors could simply absorb the structure of our proposed perceptual theorist, and therefore presumably the homunculus, into one enormous observer. Let us refer to such an observer as the *homunculus observer*. The critical question is: what, if anything, would be lost in such a change to an observer framework?

Many of the structural elements of our perceiver/theorist would be lost within a homunculus observer formulation. One loss is the ability to make certain assumptions conditional on the *solutions* of other problems that the system may pose. For example, if it can be shown that the ball in Chapter 4 could be stably supported, then one could assume there was a mode for the ball to be at rest. To determine stable support one can check if a particular linear programming problem has a solution [4]. Such a notion of stable support would constitute a theory fragment in our proposed perceptual theorist. But what would this look like for the homunculus observer? Without a decomposition of this observer, the notion of stable support would show up as structure within the prior distribution μ. Here the prior distribution μ is over all possible scenes (not just all possible scenes for a given image, but *all* possible scenes). For scenes in which an object is being stably supported by another, an omniscient onlooker might notice the regularity that there is also a mode within μ for that object to be at rest. Essentially, the homunculus observer might as well be 'doing the stability test' by table look-up in a table of all possible scenes, a table that might well be infinite. Our point here is simply that some regularities within our world are most efficiently represented in terms of whether or not a particular set of constraints has a solution (or by the character of that solution), and that there may not be a convenient bottom-up way to describe the same regularity.

However, we believe the critical loss is that, without further constraints, the homunculus observer would be unable to learn from visual experience or, equivalently, to draw inductive generalizations from examples. By "learning" here we mean not only the development of new domain theories over a lifetime of observations, but also the generation of new inductive hypotheses during the interpretation of a single scene. The argument is that, in order to learn, the homunculus observer needs to adjust its prior, μ, on the basis of visual data. However, during the course of its lifetime, due to the bandlimited nature of its transducers, it would receive data confined essentially to a finite dimensional set. Given that the prior μ is restricted only to be a probability measure on some infinite dimensional set encompassing all possible scenes, such observational data would provide an insignificant constraint on the choice of μ. The problem in essence is that the perceiver must choose from an infinite space of possible hypotheses on the basis of a data set which is invariably finite (and potentially quite small). Something extra would be needed to bias the homunculus observer to generalize appropriately, for example, from a finite number of samples of stably supported objects to a more general notion of stable support. Thus, one of the primary motivations for considering a perceptual system to be a theorist in its own right is to have it naturally incorporate useful biases for rapid and efficient learning [5].

Once the view of the perceptual system as a theorist is taken, then, the critical research questions are what is the overall form of the theories and what are the constraints or biases on the set of available theories. Are there useful general-purpose constraints and biases on the overall form of domain theories such that the theorist can rapidly adapt and learn useful theories about its world? This question can be studied two ways. The first is to study the structure in our world, with the goal being to capture useful world models within a narrowly defined set of domain theories. The second aspect is to study biological perceivers with regards to what sort of domain theories they rapidly learn. We discuss these two aspects in the following two sections.

3 Defining the perceptual theorist

In order for the notion of the perceptual theorist to be valuable it must be the case that our world is sufficiently structured so that some fairly general purpose biases on the set of available theories are both useful and applicable. Our critique of the homunculus observer rests on just such an assumption about world structure. It is therefore critical to our argument that we indicate that there may indeed be ways to provide suitable general purpose biases. To do this we briefly discuss some recent results on one way in which effective biases might be built into a perceiver/theorist, namely by constraining the very form of the conceptualization, along with the forms of the domain theories to be considered.

The basic idea is that the perceiver/theorist includes a built-in bias for mechanically building domain theories (i.e., world models) in terms of a particular set of categories, which we refer to as the 'latent categorization.' The critical point is that not *all* possible categories are in this latent set, and thus not all subsets can be entertained as hypotheses by the perceiver/theorist. Rather, only certain ones are selected according to this built-in bias (see also [6]). Moreover, we assume that the perceiver/theorist can form qualitative probabilistic domain theories of the general form considered in Chapter 4, but with the constraint that each mode in the prior distribution must correspond to some category within the latent set. Therefore this specification corresponds to a required reduction in inductive ambiguity and, as we discuss below, makes effective learning and hypothesis formation possible.

For a specific example consider the ball-in-a-box domain described in Chapter 4. An appropriate latent categorization can be constructed from the elementary concepts of smooth motion, being at rest within a ground-based reference frame, abrupt changes in velocity, contact relations between parts or objects, surface normals and the direction of gravity. The idea is that some representation of these basic concepts is built into our novice perceiver, or

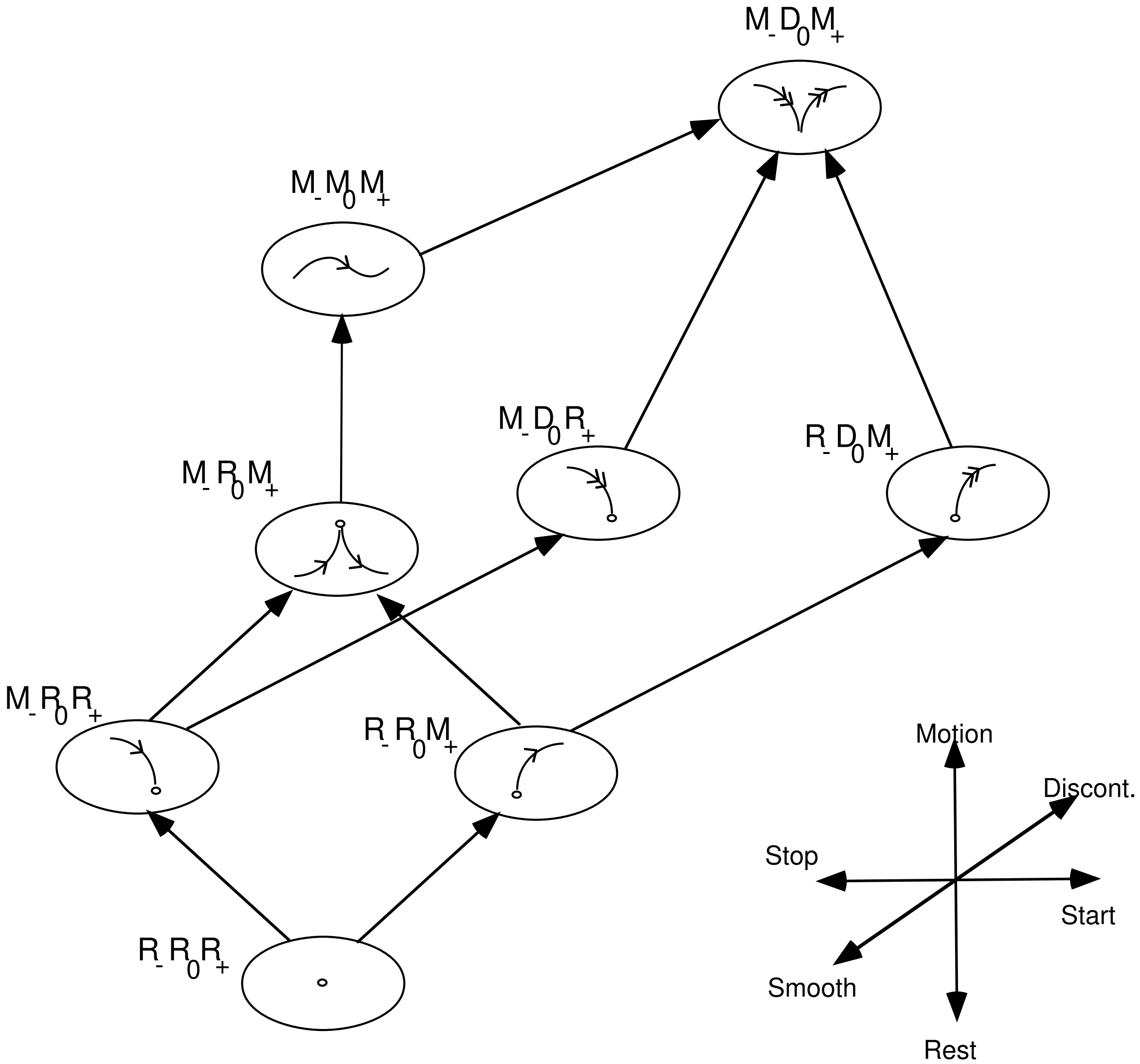

Figure 1: The structure ordering for the motion categories arising from the basic distinctions 'move' versus 'at-rest,' and 'smooth' versus 'discontinuous' motion. Here τ_- and τ_+ are two consecutive open temporal intervals separated by the instant t_0. The notation M_0, R_0, D_0 denote smooth motion, at rest, and a velocity discontinuity, respectively, at the instant t_0. Similarly, the subscripts '$-$' and '$+$' refer to τ_- and τ_+ in place of t_0.

at least is available at the stage of development in which the perceiver is ready to learn the required models. Moreover, we assume the perceiver is provided with a language which can express combinations of these basic concepts along with constraints on allowable combinations. An example of such a constraint is that an object's trajectory must be continuous. The latent set of categories is then defined to be the set of all ways to compose the basic concepts which satisfy the given constraints (see [3]). Not all of these categories are expected to be explicitly generated, nor are they all expected to turn out to be useful, but rather only some subset of them are expected to be used within domain theories.

A concrete example of a portion of such a latent set is given in Figure 1, where we have used only the two primitive distinctions: 'move' versus 'at rest,' and 'smooth' versus 'discontinuous' motion. These primitives are combined subject to the constraints: 1) there are only finitely many velocity discontinuities in any finite time interval; 2) the trajectory is continuous (but the velocity may not be); and 3) the velocity at a discontinuity is not defined. The result is that only the eight categories depicted in Figure 1 satisfy these constraints locally about any instant t_0. Note that we can split these eight categories into six different types of motion events occuring at t_0 (eg. a pause $M_- R_0 M_+$), along with two non-events (i.e., continuing to move smoothly or remaining at rest). It is encouraging to note that Rubin [7] has previously identified each of these states as perceptually salient.

This approach for generating the latent set of categories also places a partial order on the degree of generalization entailed by the various categories (i.e., the 'strength' of each inductive hypothesis). This 'structure ordering' for our eight motion categories is also depicted in Figure 1, with more constrained (and hence more structured) categories appearing lower in the figure. In particular, lower categories can be considered to be subsets of each category above it. For example, a pause event is considered to be more structured than a smooth motion event, since the pause event has the additional constraint that the velocity must be zero at t_0. Similarly, trajectories which have a velocity discontinuity at t_0 (but are smooth in some adjacent open intervals $\tau_\pm$) are the least structured events across t_0, while staying at rest across t_0 is the most structured event.

This latent set and structure ordering can function as an important bias in theory formation. Such a bias can arise, for example, by constraining the available domain theories to take the form of qualitative mixture models where *each mode corresponds to one category* within the latent categorization. Indeed, to take a specific case, the motion of a ball in the air is described by only one of the eight categories in Figure 1, namely the smooth motion era. The appropriate mixture model therefore involves a single component containing only the mode for smooth motion, as desired. Alternatively, consider a model for the motion of a house-fly. We assume our perceptual system cannot distinguish the sharp turns of a fly from a velocity discontinuity. As a result, the motion can be represented with two modes in a qualitative probability distribution, one for smooth motion and one for velocity discontinuities. This simple characterization provides a lot of information about the motion of a fly, namely they do not typically hover, stop, or start while in mid air. Moreover, the basic idea of selecting particular categories and treating them as modes in the qualitative prior distribution also naturally leads to the hypothesis that the discontinuities in a fly's motion appear in some smooth (but unknown) distribution over free space and over time. Such a hypothesis can be uniquely selected given just a single observation of a fly executing an apparent velocity discontinuity in mid-air!

Our proposed constraint on the domain theories, namely that the various modes must be specified in terms of particular categories within the latent set, limits the theorist to a discrete finite set of domain theories. This therefore represents an extremely strong bias on the prior distributions that are to be considered, which we argued above is essential for rapidly learning and revising domain theories. Indeed, because of this bias we conjecture that it is possible, for example, for a novice perceiver constructed along these lines to rapidly learn the modal structure of the ball-in-a-box theory presented in Chapter 4.

4 Perceptual theorists and psychophysics

Finally, we note that such specific models of theory formation by the perceiver/theorist, as sketched in the previous section, can be used to generate extremely specific predictions for psychophysical experiments. Following the arguments above, such predictions might be generated most readily with respect to how human observers draw inductive generalizations about novel domains. That is, we need to give the perceiver/theorist a task which exercises it's capabilities for learning or revising theories. In fact, a set of such experiments has already been carried out (see [3]). The results strongly indicate that human observers can infer the sort of mixture model described above, given just one example. Admittedly this is just a single experiment but, given our biased view of perceivers as entities which build and manipulate qualitative probabilistic theories, we are confident that the reader can make a reliable inference about the nature of perception from just this single example. (For further confirmation, see also [2, 1].)

References

[1] J. Feldman. Regular modes in inductive categories. In preparation.

[2] J. Feldman. Regularity based perceptual grouping. In preparation.

[3] J. Feldman. *Perceptual Categories and World Regularities*. PhD thesis, Department of Brain and Cognitive Sciences, MIT, 1992.

[4] B. Neumann M. Blum, A. Griffith. A stability test for configurations of blocks. Memo 188, MIT AI Lab., Feb. 1970.

[5] T. M. Mitchell. The need for biases in learning generalizations. Tech. Rep. CBM-TR 117, Rutgers University, 1980.

[6] N. Moray. A lattice theory approach to the structure of mental models. *Phil. Trans. R. Soc. Lond. B*, 327:577–583, 1990.

[7] J. Rubin. *Categories of visual motion.* PhD thesis, Department of Brain and Cognitive Sciences, MIT, 1986.

[8] A. Yuille and J. Clarke. Bayesian models, deformable templates and competitive priors. In L. Harris and M. Jenkin, editors, *Spatial Vision in Humans and Robots.* Cambridge Univ. Press, 1993.