Random Switching and Optimal Processing in the Perception of Ambiguous Signals

William Bialek[1] and Michael DeWeese[1,2]
[1]*NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540*
[2]*Department of Physics, Princeton University, Princeton, New Jersey 08544*
(Received 1 December 1994)

The optimal interpretation of noisy data is a compromise between the data and our prior expectations. In the case of ambiguous signals, like the Necker cube, trajectories of the optimal interpretation map to configurations of a random field Ising model. This analogy provides at least a qualitative account of several robust phenomena in human perception of ambiguous stimuli.

PACS numbers: 87.10.+e

When we view ambiguous figures, such as the Necker cube [1], we have the subjective impression that our perception switches at random between two equally plausible interpretations of the input data. This impression of randomness has been quantified in psychophysical experiments [2]. Clearly the different metastable percepts correspond to different patterns of neural activity in the brain, and one possible model of the random switching is that it is driven by noise in the neural circuitry itself [3]. The idea of trapping in locally stable states of neural activity could also explain perceptual hysteresis. Common to these models is that our perceptions are limited by the quality of neural hardware—the brain is behaving randomly when the external world is static. Here we present an alternative view: The brain always finds the statistically optimal interpretation of the incoming sense data, but these data are noisy and must be smoothed by some *a priori* hypotheses about the dynamics of the world [4].

We consider for simplicity that the interpretation of an ambiguous stimulus can be reduced to the problem of estimating a single variable or feature which may vary in time, $f(t)$. To arrive at this estimate the brain makes use of some sense data which we can collect into an array $d(t)$. Given these data, what can the brain (or any machine, for that matter) conclude about $f(t)$? Since the data are noisy, all one can state is the relative likelihood that the data were generated by different features, which is the conditional probability of $f(t)$ given $d(t)$. Bayes' theorem tells us that this probability can be written as

$$P[f(t)\,|\,d(t)] = \frac{1}{P[d(t)]} P[d(t)\,|\,f(t)] P[f(t)], \quad (1)$$

where $P[f(t)]$ is the *a priori* probability of the time variation $f(t)$. This prior distribution embodies the observer's knowledge that rapid variations in the feature f are unlikely in the natural world or in a given experimental setup. The distribution of the data averaged over all trajectories of the feature $f(t)$, $P[d(t)]$, serves as a normalization factor and will play no role in our general discussion.

To proceed we need to make some hypotheses about the structure of the distributions in Eq. (1). If we look at one instant of time, we can define some notion of "goodness of fit" between the data d and some possible value of the feature f. We will call this goodness of fit $\chi^2[d(t); f(t)]$, since under sufficiently strong assumptions the conventional χ^2 statistic is the relevant measure. For simplicity we assume that the fluctuations in the data are effectively white noise, so that we can write

$$P[d(t)\,|\,f(t)] \propto \exp\left(-\frac{1}{2N}\int dt\, \chi^2[d(t); f(t)]\right), \quad (2)$$

where N is the noise level. The fact that we are viewing an ambiguous figure means that there are two distinct values of f, separated by a difference Δf, which minimize χ^2. In particular, if we ignore the noise in d and set $d(t) = \bar{d}$, then $\chi^2[\bar{d}; f(t)]$ has two degenerate minima so that the two alternative interpretations of the data are equally likely. If we try to change f continuously from one stable interpretation to the other, we must surmount a barrier in χ^2, and we refer to the height of this barrier as $\chi^2_{\max}$.

To summarize our knowledge that features vary slowly, we assume that the time derivative of f is chosen independently at each instant of time from a Gaussian distribution. This means that our *a priori* distribution corresponds to a random walk of the feature, with effective diffusion constant D,

$$P[f(t)] \propto \exp\left(-\frac{1}{4D}\int dt\, \dot{f}^2(t)\right). \quad (3)$$

Putting these terms together and assuming that the noise $\delta d(t)$ in the data is small, we have

$$P[f(t)\,|\,d(t)] \propto \exp\left(-\frac{1}{4D}\int dt\, \dot{f}^2(t) - \frac{1}{2N}\int dt\, \chi^2[\bar{d}(t); f(t)]\right) \exp\left(-\frac{1}{2N}\int dt\, \frac{\partial \chi^2[d(t); f(t)]}{\partial d(t)}\bigg|_{d(t)=\bar{d}} \delta d(t)\right). \quad (4)$$

If we leave aside (for the moment) the term involving the noise δd, the probability distribution for $f(t)$ is exactly the (imaginary time) path integral for a quantum-mechanical particle moving in a double potential well. Obviously the most likely trajectory $f(t)$ is a constant value which sits at one or the other minimum of χ^2. In the present context this means that the most likely estimate of the feature f is one which is constant at one or the other of the two possible stable interpretations of the ambiguous figure. But from the quantum-mechanical analogy we know that there are also instanton trajectories which "tunnel" from one interpretation to the other [5]. These switching events occur in a small time τ_0, and in the absence of any other perturbations the mean time between switching events is $\tau_{\rm switch} = \tau_0 \exp(S_0)$, where

$$\tau_0 \sim |\Delta f| \sqrt{\frac{N}{2D\chi^2_{\max}}}\,, \qquad S_0 \sim \frac{|\Delta f|}{2}\sqrt{\frac{\chi^2_{\max}}{2DN}}\,. \tag{5}$$

How is this changed by the perturbation due to the noise $\delta d(t)$?

If we ignore dynamics on times faster than τ_0, the particle can be in only one of two states, which we identify with the states ± 1 of an Ising spin σ_n; the index n counts time in bins of size τ_0. The fact that flips are rare means that spins in adjacent time bins tend to be parallel. Small white noise fluctuations in the data can favor either stable interpretation, producing an equivalent magnetic field h_n which is an independent random variable in each bin. We emphasize that our estimate of the feature $f(t)$ is now approximated as the sequence of spins, while the random field arises from a particular instance of noise in the input data; thus the random field must be viewed as quenched disorder. The probability for a configuration of spins $\{\sigma_n\}$ is then given by the one-dimensional random field Ising model,

$$P[\{\sigma_n\}] \propto \exp\left[S_0 \sum_n \sigma_n \sigma_{n+1} + \sum_n h_n \sigma_n\right], \tag{6}$$

where the random field has a variance

$$\langle h_n^2 \rangle = \frac{\tau_0}{4N}\left(\frac{\partial \Delta\chi^2[d(t); f(t)]}{\partial d(t)}\bigg|_{d(t)=\bar{d}}\right)^2, \tag{7}$$

$$\Delta\chi^2[d(t); f(t)] = \chi^2[d(t); f(t)]|_{f(t)=f+} - \chi^2[d(t); f(t)]|_{f(t)=f-}\,. \tag{8}$$

The qualitative solution of the random field Ising model is given by the Imry-Ma [6] argument, which tells us that at low noise level (so that both S_0 and $\langle h_n^2\rangle$ are large) the configuration of spins breaks into domains of spin up and spin down, with the typical domain size

$$\xi = S_0^2/\langle h_n^2\rangle\,. \tag{9}$$

In the present context this means that typical estimates of the stimulus parameters will flip between the two stable configurations with a typical switching time $\tilde{\tau}_{\rm switch} \sim \xi\tau_0$ rather than $\tau_{\rm switch}$ from above. Putting the various factors together we find

$$\tilde{\tau}_{\rm switch} \sim \frac{2|\Delta f|^2}{D}\chi^2_{\max}\left(\frac{\partial \Delta\chi^2[d(t); f(t)]}{\partial d(t)}\bigg|_{d(t)=\bar{d}}\right)^{-2}. \tag{10}$$

At low noise levels, the optimal interpretation of ambiguous incoming data thus switches randomly at a rate *independent* of the noise level. Indeed, the switching rate is proportional to the *a priori* expected drift rate between the two ambiguous interpretations, $|\Delta f|^2/D$, but is suppressed in (linear) proportion to the χ^2 "barrier" between the interpretations.

These results should be contrasted with models where random switching among percepts is triggered by dynamics and noise in a neural network rather than by the noise in the sense data itself. In general, the switching between locally stable states of the network must be some form of Kramers' problem [7], and we expect that the predicted switching rate will depend exponentially on the noise level and on the "barrier" height. One might object that our analysis is based on a probability distribution for trajectories $f(t)$, and it ought to be possible to construct a noisy network whose trajectories are drawn from the same distribution. This is true, but the resulting network dynamics are quite complex and nonlocal in time; this is related to the fact that the real devices which make optimal estimates usually make delayed estimates [8]. The central point of our discussion is that random switching is an inevitable feature of optimal estimation, independent of the circuitry in which the estimator is realized.

To make these ideas concrete we consider a problem in pitch perception [9]. When we hear a harmonic sequence, e.g., 1000, 1200, and 1400 Hz, we assign a pitch equal to the fundamental even if it is not present in the physical signal. This search for the missing fundamental continues if the signals are slightly inharmonic, as in a sequence $f_n = nf_1 + \delta$, and the perceived pitches can be predicted [9] as those f which minimize

$$\chi^2 = \sum_\mu (f_\mu - n_\mu f)^2. \tag{11}$$

Note that there are multiple minima corresponding to different assignments of the integers $\{n_\mu\}$. When the signals are maximally inharmonic, $\delta = f_1/2$ and there are two near degenerate minima of χ^2; if we choose the set $\{n_\mu\}$ which minimizes χ^2 at fixed f we find that the resulting $\chi^2(f)$ has the standard double-well form. Human observers hear both pitches, and the percept switches at random, as for ambiguous figures in vision.

In the context of our analysis above, the parameter we are trying to estimate is the pitch f, and the data are the representations of the individual components f_μ in the sensory nerves. A typical experiment involves listening to three components which are interpretable as the 10th, 11th, and 12th (or 9th, 10th, and 11th) harmonics of a fundamental near 200 Hz; the two possible interpretations differ by $|\Delta f| \sim 20$ Hz. In naturally occurring sounds, such as speech, frequency modulations of this magnitude occur on time scales of several tens of msec, so we human listeners probably have an *a priori* distribution for pitch fluctuations with $|\Delta f|^2/2D \sim 0.01$–$0.1$ sec. The remaining factor in Eq. (9) is dimensionless and determined entirely by the choice of integers $\{n_\mu\}$, and we find the predicted $\tau_{\text{switch}} \sim 1.2-12$ sec, in reasonable agreement with experiment [2,9,10].

The predicted switching times depend on the parameters of the stimulus *and* on the *a priori* assumptions of the observer. Thus individual differences in mean switching time could be substantial, although the distribution of switching times should be more reproducible, in agreement with experiment [2]. If the brain adjusts its "prior" expectations in response to recent sensory experience [4], it should be possible to manipulate the observable switching times by having the observer listen to sounds with different statistical properties. Certainly human observers adjust their expectations in relation to the instructions given at the start of the experiment, so that instructing the observer to expect changing signals should increase the reversal rate, as observed [11].

In vision one can have not just one ambiguous figure but a whole array, in one or two spatial dimensions. Ramachandran and Anstis [12] have studied arrays of alternating dots which can be seen as moving either horizontally or vertically. In this case the feature $f(t)$ is the orientation of the local motion vector, and it is clear that a reasonable *a priori* distribution will include not only the penalties for temporal variation discussed above but also a term which penalizes spatial gradients of the velocity. In this case the same arguments given here map the problem of making optimal estimates for the array onto a two- or three-dimensional random field Ising model. At zero noise, of course, both models have a true phase transition to ferromagnetic order, and this ordering survives the random field perturbation in the 3D case; the 2D case is marginal [6]. A ferromagnetic phase corresponds to perceiving all of the dot pairs moving in the same direction, or a coherent motion percept. For two-dimensional arrays, then, it is clear that one should observe coherent motion, and switching between the two possible directions of motion should be very slow, infinitely slow in the limit of large arrays.

Experiments [12] show clearly that a 2D array is always seen to move coherently, and spontaneous switching is unobservably slow, as predicted. This clear "magnetization" of the pattern is disrupted if a stripe through the 2D array is occluded, as expected since the corresponding couplings in the Ising model are eliminated. In an array of Necker cubes there is no obvious *a priori* distribution which would couple the relative depth parameters in neighboring cubes, so we expect a paramagnetic phase where different cubes switch at random, as observed [13].

A coherent percept across the entire 2D array is predicted to occur even when the prior distribution has only local terms, and hence the optimal processor can be constructed entirely from local operations. True "magnetization" in the 1D array would require long range interactions, but it may be difficult to distinguish this from the marginal case. Controlled manipulation of the noise level, perhaps by jittering the dots which give rise to ambiguous motion, might make it possible to test the prediction that the transition to coherent motion is indeed a phase transition. One can even fantasize about studying the scaling behavior of our perceptions in the neighborhood of such a transition, but this may be asking too much of the experiments.

To summarize, we have shown that optimal estimation of a *potentially* time-varying feature inevitably leads to random perceptual switching in response to ambiguous signals. The predicted switching rate is independent of the (small) noise level in the sense data, but does depend on the observer's *a priori* hypotheses. If the observer can assume that features are truly static ($D \rightarrow 0$), then the predicted switching rate vanishes, in agreement with experiments on carefully instructed human observers [11]. It is attractive that this combination of randomness and apparent subjectivity emerges from an objective theory of optimal estimation.

We thank R. Mannella for organizing the conference which made us think about these issues, and S. Smirnakis for helpful discussions. M. Elbaum, M. Feigenbaum, and M. Magnasco all asked the same important question [8], and we are grateful to them for insisting on an answer.

[1] L. A. Necker, Phil. Mag. Ser. 3 **1**, 329 (1832). For a brief historical review see G. H. Fisher, Perception & Psychophysics **4**, 189 (1968).

[2] A. Borsellino, A. De Marco, A. Allazetta, S. Rinsei, and B. Bartolini, Kybernetik **10**, 139 (1972).

[3] See, for example, T. Dtzinger and H. Haken, Biol Cybernet. **63**, 453 (1990); F. Masulli, M. Riani, and E. Simonotto, Proc. SPIE Int. Soc. Opt. Eng. **1710**, 267 (1992); M. Riani and E. Simonotto, Phys. Rev. Lett. **72**, 3120 (1994).

[4] Evidence that the brain can perform optimal processing of sensory signals comes from many systems, ranging from the counting of single photons in dark-adapted vision (both humans and toads), to the estimation of motion in the fly visual system and the discrimination of target range in bat echolocation. For general reviews see W. Bialek, in *Princeton Lectures on Biophysics* (World Scientific, Singapore, 1992), pp. 321-401, and H. B. Barlow, Proc. R. Soc. London Sect. B **212**, 1 (1981). For the importance of *a priori* hypotheses in determining the form of the optimal estimator, see M. Potters and W. Bialek J. Phys. I (France) **4**, 1755 (1994).

[5] S. Coleman, *Aspects of Symmetry* (Cambridge University Press, Cambridge, 1975).

[6] Y. Imry and S. K. Ma, Phys. Rev. Lett. **35**, 1399 (1975).

[7] H. A. Kramers, Physica (Utrecht) **7**, 284 (1940).

[8] The question here is whether there exists a classical stochastic system with trajectories that have the same statistics as the imaginary time trajectories of a given quantum system acted upon by a time-dependent external force. The answer is yes, but even for the harmonic oscillator one can show that corresponding classical force is a nonlocal functional of the force in the quantum

problem. The harmonic oscillator problem is analogous to optimal estimation of a signal with Lorentzian spectrum in a background of Gaussian white noise, and the nonlocality of the classical force is related to the Wiener-Kolmogorov theory of optimal causal filters in the estimation problem [N. Wiener, *Extrapolation, Interpolation and Smoothing of Time Series* (Wiley, New York, 1949)].

[9] For a review see E. de Boer, Handbk. Sens. Physiol. **V/3**, 479 (1976).

[10] For the precise form of χ^2 in Eq. (11), the action and width of the single instanton differ slightly from the estimated in Eq. (5). Carrying through the more exact calculation gives a factor of 2 reduction in the estimate of $\tilde{\tau}_{\mathrm{switch}}$, well within the uncertainties of D.

[11] J. J. Girgus, I. Rock, and R. Egatz, Perception & Psychophysics **22**, 550 (1977); I. Rock and K. Mitchener, Perception **21**, 39 (1992).

[12] V. A. Ramachandran and S. M. Anstis, Perception **14**, 135 (1985).

[13] G. M. Long and T. C. Toppino, Perception **10**, 331 (1981).