Correlating Spikes and Behavior

Theoretical Neuroscience

Fabrizio Gabbiani
Division of Neuroscience
Baylor College of Medicine
One Baylor Plaza
Houston, TX 77030
e-mail:gabbiani@bcm.tmc.edu
1 Introduction

Studying how sensory perception arises from the encoding and processing of information by nerve cells and neuronal networks is probably one of the most fascinating and challenging aspect of systems neuroscience. The sensory stimuli to which animal species respond and the behaviors that they elicit are so diverse that a multitude of approaches and techniques have been devoted to this goal. In this lecture, we will focus on a very restricted set of sensory perception tasks involving the detection of signals embedded in noise. These tasks have been studied at the level of individual human subjects, a field called psychophysics. Many of the methods used in psychophysics are closely related to those originally developed for signal detection in an engineering context. We will see how these methods can also be applied to study perception in animals and to analyze neuronal signals, thus opening a way to relate perception to neuronal processing.

2 Single photon detection in the visual system

The Hecht, Shlaer and Pirenne (HSP) experiment. In 1942, Hecht, Shlaer and Pirenne published a study in which they investigated the threshold of human subjects for detecting brief, weak light flashes. The experimental conditions were carefully optimized to maximize the sensitivity of the human subjects. Prior to the task, subjects were kept in the dark for at least 30 mins to ensure full dark-adaptation. The flashes were delivered at a horizontal distance 20 degrees away from the fovea in a region where the density of rod photoreceptors is high. The area covered by the stimulus (10 minutes of arc) was also optimized to yield the highest sensitivity. Stimuli were presented for 1 msec and the wavelength of the light stimulus was 510 µm (green), a value at which the eye is known to be most sensitive for dim vision. In the experiments, the energy of the light flash or equivalently the mean number of photons delivered at the cornea was varied and the frequency at which the observers detected the flashes was recorded.

The results of the experiments are presented in Fig. 1 for 3 subjects. Typically, the number of photons at the cornea needed to detect 60 percent of the flashes ranged between an average of 54 and 148 light quanta. Based on the data available at the time, Hecht Shlaer and Pirenne estimated that 4 percent of the light would be reflected by the cornea, 50 percent of the remaining photons would be absorbed by the ocular media before reaching the retina and 80 percent of the light would pass through the retina without being absorbed by photoreceptors. Thus, only about 9.6 percent of the photons available at the cornea could be responsible for light detection in these experiments. This corresponds to an average of 5-14 light quanta. This number is surprisingly small and suggests that absorption of two photons by the same photoreceptor is highly improbable since the area covered by the light stimulus corresponded to approximately 500 rods photoreceptors (4 percent probability). Thus, one predicts that rods should be sensitive to single photons and that the simultaneous absorption of a small number of them leads
to conscious sensation.

Because the average number of absorbed photons is so small, one expects considerable fluctuations in the number of photons absorbed from trial to trial. Thus, it is conceivable that a large fraction of the subject response variability is caused by fluctuations in the absorbed photon number. If we assume that photon absorptions are independent random events of constant probability, we expect their distribution to follow a Poisson distribution, just as the number of photons emitted by the light source and observed at the cornea. Let \( a \) be the average number of absorbed photons for a given average flash intensity. Hecht, Shlaer and Pirenne assumed that \( a = \alpha n \), where \( n \) is the average number of photons measured at the cornea and \( \alpha \) is an attenuation factor related to the optical properties of the eye and retina. Let \( P(k) \) denotes the probability of \( k \) photons being absorbed, then

\[
P(k) = \frac{a^k}{k!} e^{-a}.
\]

If a human observer sees the experimental light flash only when a fixed threshold number of photons \( k_0 \) is absorbed, we expect a probability of seeing the stimulus given by

\[
P_D(a) = \sum_{k \geq k_0} \frac{a^k}{k!} e^{-a}.
\]

The curves \( P_D \) are plotted as a function of \( \log_{10}(a) \) for various values of \( k_0 \) in Fig. 2a. The average number of absorbed photons \( a \) for a given average number of corneal photons \( n \) is of course unknown. If the probability of seeing \( P_D \) is plotted as a function of \( \log_{10}(n) \), the curve becomes identical in shape to that determined by \( P_D \) as a function of \( \log_{10}(a) \) except for a shift along the horizontal axis, since \( \log_{10}(a) = \log_{10}(n) + \log_{10}(\alpha) \). Fitting the appropriate value of \( k_0 \) to the experimental data then becomes very easy: it simply amounts to matching the curve’s shape to that of the cumulative Poisson distributions of eqs. (1). Thus, the two parameters of the model, \( k_0 \) and \( \alpha \), are determined by the slope of the frequency of seeing curve and its shift along the abscissa, respectively. The fits obtained in Fig. 1 for the probability of seeing as a function of the average number of corneal photons matches well this expectation for values of \( k_0 \) between 5 and 7.

Barlow’s dark light hypothesis. The HSP experiment suggests that most of the variability in the observers’ responses is due to noise in the physical stimulus rather than biological noise. As pointed out a decade later, the experimental design and its interpretation have however several shortcomings:

1. If rods are indeed sensitive to single photons, why would observers not be as well, given that biological noise is assumed to be inexistant?

2. The HSP experiment is by itself somewhat ambiguous: an observer could always lower its threshold and thus give the appearance of “seing” better.

Barlow proposed a solution to these two problems by interpreting the results differently and by proposing a modified model of photon absorption. Following Hecht, Shlaer and
Pirenne he proposed that rods are sensitive to single photons, but he postulated that several rods must be activated simultaneously when a weak flash is detected to overcome biological noise. One plausible source of noise is the random spontaneous decay of the rod photopigments (rhodopsin) in the absence of light. This decay would give the illusion of photon arrival and thus the registration of a single photon would in turn be unreliable to signal the presence of weak light flashes. Other sources of noise might result from central nervous system processing and were lumped together with spontaneous rhodopsin decay in the Barlow model.

Let us assume that in the absence of light the mean number of absorbed photons (dark light) is $x$ and follows a Poisson distribution. When presented with "blank" trials where no flash occurs, an observer is expected to report a light flash (even if none occurred) in a fraction of the trials because of the noise. If we call $P_{FA}$ the probability of such "false-alarms", it is given by

$$P_{FA}(x) = \sum_{k=k_0}^{\infty} \frac{x^k}{k!} e^{-x}. \quad (2)$$

It depends both on the amount of noise ($x$) and the detection threshold ($k_0$) of the observer. In the presence of a light flash, the mean number of absorbed photons will be due both to absorption related to the light flash, $\alpha n$, and to the noise, $x$. If both processes follow independent Poisson distributions, their sum is also Poisson with mean $a = \alpha n + x$. Thus,

$$P_D(a) = \sum_{k=k_0}^{\infty} \frac{(\alpha n + x)^k}{k!} e^{-(\alpha n + x)}. \quad (3)$$

The model has three parameters (instead of two in the HSP formulation): the threshold level, $k_0$, the fraction of absorbed photons, $\alpha$, and the "dark light level" $x$. Fitting the model to the data now becomes more complex because the parameters cannot be simply interpreted geometrically. The additional parameter can be fit to the data by using the false-alarm rate obtained from presenting "blank" trials. As illustrated in Fig. 2, the model offers good fits to the HSP data. Typically, the fraction of absorbed photons $\alpha$ is predicted to be higher in the presence of noise $x \neq 0$, this is consistent with later estimates of the probability of photon absorption (predicted to be higher, $\sim 20\%$ than at the time of the HSP experiment). Furthermore, by encouraging subjects to report less probable stimuli, the threshold is observed to decrease in parallel with an increase in the probability of false-alarms. This is in agreement with point 2 above and emphasizes the needs to monitor thresholds with independent data.

**Detection of light in dark-adapted retinal ganglion cells.** How does the performance of neurons in detecting weak light flashes compare with the observer’s performance? Since retinal ganglion cells are the first spiking neurons that convey information to the central nervous system, it is natural to investigate their responses to such weak light flashes. The experiments were performed in the cat using ON-center retinal ganglion cells and a representative experimental result is illustrated in Fig. 3. The stimulus consisted
either of a weak light flash (5 photons on average) of 10 ms duration or of a "blank" trial. The spiking response of the retinal ganglion cell was recorded during a time window of 200 ms starting at flash onset. In the absence of light, the cell was spontaneously active with an average of 4.14 spk whereas in the presence of light the mean spike count was increased to 6.62. Does the distribution of spike counts match the Barlow model described above? If this were the case, one would expect the spike counts to be Poisson distributed both for the spontaneous and evoked response with a difference in means equal to the mean number of absorbed photons, \( \Delta m = q_a = \alpha n \) and a difference in variance \( \Delta \sigma^2 = q_a \) so that the Fano factor given by \( \Delta \sigma^2/\Delta m = 1 \). However, the experimentally measured difference in variance is usually larger than that expected from a Poisson distribution. Let us assume that for each absorbed photon an average of \( \lambda \) spikes are produced. Then \( \Delta m = \lambda q_a \) and \( \Delta \sigma^2 = \lambda^2 q_a \) so that \( \Delta \sigma^2/\Delta m = \lambda \). The variance in the evoked spike count distributions is consistent with the assumption that between 2 and 3 spikes are fired in response to each absorbed photon (i.e., \( 2 \leq \lambda \leq 3 \)). Thus, the response of retinal ganglion cells is consistant with a process of amplification of the absorbed photons at low light levels.

The performance of retinal ganglion cells at detecting light can be assessed by chosing a fixed threshold spike count \( \lambda_{thres} \) and computing the corresponding probability of detecting the light flash in the above experiment. In a trial that consists with equal probability of a light flash or a "blank", the observer will report that a light flash occurred if \( \lambda_{thres} \) or more spikes are counted. Otherwise the observer reports that no flash occurred ("blank" trial). As pointed out above, the probability of detection, \( P_D = P(\lambda \geq \lambda_{thres}|\text{flash}) \) will of course depend on the selected threshold: decreasing \( \lambda_{thres} \) leads to higher probabilities of detection. This is however offset by an increase in the probability of false-alarms, \( P_{FA} = P(\lambda \geq \lambda_{thres}|\text{blank}) \), i.e., the probability of reporting a flash in a "blank" trial. A plot of \( P_D \) as a function of \( P_{FA} \) is called an ROC curve (Receiver Operating Characteristic, a term originating from early applications to radar during WWII). Such a plot is illustrated in Fig. 3C for the retinal ganglion cell of Fig. 3A, B. Each labelled dot \((1,2,3,\ldots)\) represents the performance for a spike count threshold \( \lambda_{thresh} = 1, 2, 3, \ldots \). Plotting \( P_D \) as a function of \( P_{FA} \) (instead of using directly the threshold \( \lambda_{thres} \)) is a better representation of the data because this fully characterizes the performance of the observer and is independent of the particular way in which the classification decision was made. This allows to compare performance with that of an observer based on the Barlow model: the roman numbers correspond to the \((P_{FA}, P_D)\) values obtained from the Barlow model with parameters \( \alpha = 0.18, x = 6.5 \) at detection thresholds of 1, 2 and 3 absorbed quanta, respectively. Because the Barlow model describes accurately the psychophysical performance of human observers, this suggests that the performance of single retinal ganglion cells is comparable to that of humans. This conclusion is based on the assumption that cats would report light flash occurrences in a similar manner than humans or vice-versa, that human retinal ganglion cells responds like cat retinal ganglion cells to light flashes.

Single photon sensitivity in rods. Do rods really respond to single light quanta? This question was finally answered at the beginning of the eighties when Dennis Baylor and
colleagues developed a technique that allowed to record responses of single rods isolated from the retina of salamanders to weak flashes of light. Their results unambiguously demonstrated responses to single photons, thus verifying the claim that dark-adapted rod photoreceptors are highly sensitive detection devices 40 years after the original HSP experiment.

3 Signal detection theory and psychophysics

Psychophysics is the subfield of psychology devoted to the study of physical stimuli and their interaction with sensory systems. Psychophysical tasks using weak visual stimuli or stimuli embedded in noise have been extensively used to draw conclusions on how sensory information is processed by the visual system. These tasks can be analyzed using methods originally developed in the context of engineering for the detection of weak physical signals. In this section, we introduce the formal framework used to describe and analyze the experiments reported in the previous section. This will allow to show that discrimination performed in terms of a threshold in spike number (or using the number of absorbed photon as in sect. 2) is ideal under certain circumstances. We also introduce a second type of psychophysical task, different from that considered in the previous section. This task, the 2-alternative forced choice task, will give us a better understanding of the significance of ROC curves.

Yes-no rating experiments. Experiments like those described in sect. 2 are called yes-no rating experiments. In these experiments, either one of two stimuli ($s_0$ and $s_1$) is randomly presented with equal probability. An observer is to report after each stimulus presentation which one of $s_0$ or $s_1$ was presented. In a typical situation $s_0$ is ”noise” and $s_1$ corresponds to a signal presented simultaneously with the noise (”signal plus noise”). In sect. 2 the noise condition would correspond to the ”blank” stimulus and the ”signal plus noise” condition to the flash stimulus. The responses are denoted as $r = 0$ or 1 depending on whether ”noise” or ”signal plus noise” is chosen by the observer.

Two alternative forced-choice (2-AFC) experiments. A 2-alternative forced-choice experiment is one in which the subject is required to respond only after two successive stimulus presentations. Both $s_0$ and $s_1$ are presented exactly once with equal probability in the two presentation intervals. After the second interval, the subject is asked to report in which interval $s_1$ (”signal plus noise”) was presented. In the flash detection experiments described above, this corresponds to presenting the ”blank” stimulus in one interval and the ”signal plus noise” condition to the flash stimulus. The responses are denoted as $r = 0$ or 1 indicate the first or second interval, respectively.

Correct detection and false-alarm probabilities. In a yes-no rating experiment, the probability of correct detection, $P_D$ is the probability of reporting the signal when it was
indeed present, i.e., \( P_D = P(r = 1|s_1) \) and the probability of false-alarm is the probability of incorrectly reporting the signal when it was absent, i.e., \( P_{FA} = P(r = 1|s_0) \). The total error rate of the observer is given by averaging both types of errors by their probability of occurrence,
\[
\epsilon = \frac{1}{2} P_{FA} + \frac{1}{2} (1 - P_D).
\]

In a 2-AFC experiment, we define \( P_C \) as the probability of correct response (i.e., \( r = 0 \) when \( s_1 \) was presented in the first interval and \( r = 1 \) when \( s_1 \) was presented in the second interval).

**Psychometric curves.** When the strength of the signal is continuously varied over a range of values, a plot of the detection probability as a function of signal strength is called a *psychometric curve*. Such psychometric curves can be computed either for yes-no rating or 2-AFC experiments. It is usual to define from a psychometric curve a *detection threshold* to be able to compare the responses of subjects across different conditions. Typically, detection thresholds are defined as 50% correct performance for yes-no rating experiments and 75% correct performance for 2-AFC experiments. These definitions are somewhat arbitrary and some authors define detection thresholds using different values (such as 68% correct performance for 2-AFCs).

**ROC curves.** For a yes-no rating experiment, the ROC curve is a plot of \( P_D \) as a function of \( P_{FA} \) for a fixed signal strength. In psychophysical experiments, ROC curves are often plotted for a signal strength equal to the psychophysical threshold. As explained below such ROC curves fully characterize the performance of the observer for a fixed set of physical stimulus conditions.

**Statistical distribution of stimuli or responses.** When a psychophysical detection experiment is carried out, one typically has access either to the probability distribution of the "noise" and "signal plus noise" or to some physiological variable such as the number of spikes fired by a neuron in response to "noise" and "signal plus noise". The question then arises as to how that information can be used to "optimally" decide which of the two stimuli was presented. As we will see later on, this question can be given a precise answer if we specify what "optimal" means. For concreteness, we will start by considering two examples: the first one is closely related to the model used in sect. 2 to describe photon absorption by rods. The second example is formulated in terms of stimulus noise properties instead of spike count responses. It will play an important role in the next section in a slightly different context.

**Example 1.** We assume that our response variable is Poisson distributed with mean values \( m_0 \) and \( m_1 \) (\( m_1 > m_0 \)) for stimuli \( s_0 \) and \( s_1 \). The response variable could for example represent the distributions of spikes generated by a neuron in response to the two stimuli. If \( m_0 = x \) and \( m_1 = \alpha n + x \) we obtain the distribution of activated rods to the "blank"
and flash stimuli in Barlow’s model. Thus,

\[ P(k|s_0) = \frac{m_0^k}{k!} e^{-m_0} \quad \text{and} \quad P(k|s_1) = \frac{m_1^k}{k!} e^{-m_1}. \] (4)

Let us assume that \( k \) spikes have been observed under the assumption of (4). The probabilities \( P(k|s_0/1) \) can then be thought of as the likelihood of this observation under conditions \( s_0/1 \), respectively. Thus, a natural quantity to consider is the likelihood ratio,

\[ \Lambda(k) = \frac{P(k|s_1)}{P(k|s_0)} = \left( \frac{m_1}{m_0} \right)^k e^{-(m_1-m_0)}. \]

The ratio \( \Lambda(k) \) will be large when \( k \) is much more likely to originate from \( s_1 \) than from \( s_0 \) and vice-versa. Thus, a plausible decision rule is to opt for \( s_1 \) when \( \Lambda(k) \) exceeds a threshold \( \eta \), i.e.,

\[ \Lambda(k) \geq \eta \Rightarrow s_1, \quad \Lambda(k) < \eta \Rightarrow s_0. \]

Equivalently, one may consider the threshold log \( \eta \) on the log-likelihood ratio log \( \Lambda \) since the logarithm is monotone increasing. Because log \( \Lambda = k \left( \log m_1 - \log m_0 \right) - (m_1 - m_0) \) this decision rule is equivalent to imposing a threshold on the number of spikes,

\[ k \geq k_{th} = \frac{\log(\eta) + m_1 - m_0}{\log m_1 - \log m_0} \Rightarrow s_1 \]
\[ k < k_{th} \Rightarrow s_0. \]

The probabilities of correct detection and false-alarm are given as in sect. 2 by

\[ P_D = \sum_{k \geq k_{th}} \frac{m_1^k}{k!} e^{-m_1}, \quad \text{and} \quad P_{FA} = \sum_{k \geq k_{th}} \frac{m_0^k}{k!} e^{-m_0}. \] (5)

Thus, fixing a threshold \( k_0 \) gives a probability of false-alarm \( P_{FA_0} \) and a corresponding probability of correct detection \( P_{D_0} \) as determined by eqs. (5). If \( k_{th} = k_0 + 1 \) then \( P_{FA_1} < P_{FA_0} \) and \( P_{D_1} < P_{D_0} \). What if we would like to obtain a probability of correct detection between \( P_{D_1} \) and \( P_{D_0} \), say \( \frac{1}{2}P_{D_1} + \frac{1}{2}P_{D_0} \)? This may be achieved by the following strategy: if \( k \geq k_0 + 1 \) choose \( s_1 \) and if \( k < k_0 \) choose \( s_0 \). If \( k = k_0 \) choose \( s_0 \) with probability \( \frac{1}{2} \) and \( s_1 \) with probability \( \frac{1}{2} \). This corresponds to using the decision rule determined by \( k_0 \) and the one determined by \( k_0 + 1 \) with probability \( \frac{1}{2} \) and yields a probability of correct detection that is the average of those two decision rules, i.e., \( \frac{1}{2}P_{D_1} + \frac{1}{2}P_{D_0} \). Such a decision rule is called a randomized decision rule. Although this may seem rather artificial at this point, we will see later on how this example helps understand a fundamental result on optimal decision rules. The ROC curves for such decision rules are plotted in Fig. 5.

Example 2. Let us consider a slightly different task than that considered in the previous section. A dot of fixed contrast \( c_0 \) is presented on a background whose value is drawn
from a Gaussian distribution with zero mean and standard deviation $\sigma_n$. The task is to detect the presence or absence of the dot. The distribution of the contrast "noise" $s_0 = n$ has zero mean and standard deviation $\sigma_n$:

$$p(c|s_0) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-c^2/2\sigma_n^2}.$$  

Signal and noise are assumed to add independently, yielding the distribution of $s_1 = c_0 + n$ from the value of $c_0$ and the distribution of $n$:

$$p(c|s_1) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-(c-c_0)^2/2\sigma_n^2}.$$  

For a given observed value of the contrast, $c$, we compute the log-likelihood ratio,

$$\log \Lambda(c) = \log \left( \frac{p(c|s_1)}{p(c|s_0)} \right) = \frac{c_0^2}{\sigma_n^2} \cdot c - \frac{c_0^2}{2\sigma_n^2}.$$  

Similarly as in the previous example, the log-likelihood ratio depends linearly on contrast: using a threshold value $\log \eta$ is therefore equivalent to setting a threshold on contrast to decide between $s_0$ and $s_1$. For convenience, we formulate the decision rule in terms of the normalized contrast $l = \frac{c}{\sigma_n}$ and the normalized distance $d = \frac{c_0}{\sigma_n}$ between the means of the two distributions $p(c|s_0)$ and $p(c|s_1)$:

$$l < \frac{d}{2} + \frac{\log \eta}{d} \Rightarrow s_0$$  

$$l \geq \frac{d}{2} + \frac{\log \eta}{d} \Rightarrow s_1.$$  

Because $l \sim \mathcal{N}(0,1)$ for $s_0$ and $l \sim \mathcal{N}(d,1)$ for $s_1$, the probabilities of false-alarm and correct detection are given by

$$P_{FA} = \int_{\frac{\log \eta}{d} + \frac{d}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx = \text{erfc}\left( \frac{\log \eta}{d} + \frac{d}{2} \right),$$  

$$P_D = \int_{\frac{\log \eta}{d} + \frac{d}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx = \text{erfc}\left( \frac{\log \eta}{d} - \frac{d}{2} \right).$$  

where erfc$(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx$ is the complementary error function. Thus both $P_D$ and $P_{FA}$ depend on $d$ in this example. ROC curves are plotted for different values of $d$ in Fig. 5.

**Ideal decisions rules (ideal observers).** We are now ready to define more precisely the decision rules introduced above and to state the basic result asserting that optimal (ideal) decisions are always based on the likelihood ratio.
Let $X$ be the set of values that can be taken under $s_0$ and $s_1$, irrespective of whether stimulus 0 or 1 is presented. In the first example above, $X = \mathbb{N}$ (integers) and in example 2, $X = \mathbb{R}$ (real numbers). In the context of a yes-no rating experiment, a decision rule (or equivalently a test) is a map $\phi : X \rightarrow \{0, 1\}$ assigning to each possible observation $x \in X$ either stimulus $s_0$ or stimulus $s_1$. In the context of a 2-AFC experiment a decision rule is a map $\phi : X \times X \rightarrow \{0, 1\}$ that assigns to each pair of responses $(x_1, x_2)$ a number 0 or 1 corresponding to the interval (first or second) in which the ”signal plus noise” appeared.

There are many ways of defining ideal or optimal decision rules depending on the optimality criterion chosen. We focus on the Neyman-Pearson and minimum error criteria. In the context of a yes-no rating experiment, a Neyman-Pearson ideal observer is one that maximizes the probability of detection $P_D$ for a fixed value, say $\alpha$, of the probability of false-alarm ($P_{FA}$). Such a decision rule is called most powerful test of size $\alpha > 0$. In the context of both yes-no rating and 2-AFC experiments, a minimum error ideal observer is one that minimizes the probability of error or equivalently maximizes the probability of correct decisions, $P_C$.

The fundamental result is the following

**Neyman-Pearson Lemma.** Let $P_0$ and $P_1$ be two probability distributions with densities $p_0$ and $p_1$ corresponding to two conditions $s_0$ and $s_1$. A test of the form

$$
\phi(x) = \begin{cases} 
1 & \text{if } p_1(x) > k p_0(x), \\
\gamma & \text{if } p_1(x) = k p_0(x), \\
0 & \text{if } p_1(x) < k p_0(x), 
\end{cases}
$$

for some threshold $k \geq 0$ and a number $0 \leq \gamma \leq 1$ is the most powerful test of size $\alpha > 0$. When $\phi(x) = 0$ choose $s_0$ and when $\phi(x) = 1$ choose $s_1$. If $\phi(x) = \gamma$ flip a ”$\gamma$-coin” and choose $s_1$ with probability $\gamma$ (the probability that the coin turns up heads). The test defined above is essentially unique (up to changes on a subset of values $x \in X$ with zero probability of occurrence).

The test may also be formulated in terms of the likelihood ratio, i.e.,

$$
\phi(x) = \begin{cases} 
1 & \text{if } \Lambda(x) > k, \\
\gamma & \text{if } \Lambda(x) = k, \\
0 & \text{if } \Lambda(x) < k.
\end{cases}
$$

In most cases, the probability that $\Lambda(x) = k$ is effectively zero. In example 2 above for example, both probability densities are Gaussians and thus probabilities are only non-zero over intervals of finite length. In such cases the threshold $k$ is determined by

$$
\alpha = \int_k^\infty p_\Lambda(x|s_0) \, dx, \quad (6)
$$

where $p_\Lambda(x|s_0)$ is the probability distribution of the likelihood ratio when $s_0$ is in effect.
The probability of correct detection is similarly given by

\[ P_D = \int_k^\infty p_\Lambda(x|s_1) \, dx. \]

In the case of example 1, the probability of false-alarm \( \alpha \) may lie between two values \( \alpha_0 \) and \( \alpha_1 \) determined by discrete thresholds \( k_0 \) and \( k_1 \). When this occurs, one sets \( k = k_1 \) and a randomized test is needed.

**Minimum error test.** Assume that \( s_0 \) and \( s_1 \) are presented with equal probability (1/2). The minimum error test is a likelihood ratio test with threshold \( k = 1 \). Alternatively, the minimum error test can be determined from the ROC curve by computing \( \frac{1}{2} P_{FA} + \frac{1}{2} (1 - P_D) \) as a function of \( P_{FA} \) and selecting the minimum value.

**Minimum error in a 2-AFC experiment.** If the observer’s response is not biased towards one of the two presentation intervals, the minimum error test in a 2-AFC experiment is to compare the likelihood ratio of the two presentations \((x_1, x_2)\) and select response \( r = 1 \) for the presentation interval with the highest likelihood ratio:

\[
\begin{align*}
\frac{p(x_1|s_1)}{p(x_1|s_0)} > \frac{p(x_2|s_1)}{p(x_2|s_0)} & \Rightarrow r = 0 \\
\frac{p(x_1|s_1)}{p(x_1|s_0)} < \frac{p(x_2|s_1)}{p(x_2|s_0)} & \Rightarrow r = 1
\end{align*}
\]

Note that in this case, no threshold is needed. This can be understood intuitively from the fact that one presentation interval effectively serves as the threshold for the other one.

**Properties of ROC curves.** We state without proof some of the most important geometric properties of ROC curves.

**Convexity of ROC curves.** The fact that ROC curves are convex follows by an argument similar to that used in example 1. If we have two points (tests) \((P_{FA1}, P_{D1})\) and \((P_{FA2}, P_{D2})\) on a ROC curve, the randomized tests built as linear combinations of these two tests yields a straight line connecting the two points. The most powerful tests of the Neyman-Pearson lemma have to be at least as performant as the randomized tests, i.e., they have to lie above the straight line connecting \((P_{FA1}, P_{D1})\) and \((P_{FA2}, P_{D2})\). By definition, this means that an ROC curve is convex.

**Slope of ROC curves.** The slope of an ROC curve is the threshold value of the corresponding Neyman-Pearson test. This means that

\[
\left. \frac{dP_D}{dP_{FA}} \right|_\alpha = k,
\]

where \( k \) is determined by eq. (6).
Area under an ROC curve. The area under an ROC curve for a yes-no rating task equals the expected ideal observer performance in the corresponding 2-AFC task. The area under an ROC curve is thus often used as a measure of discrimination performance, since it is independent of the chosen threshold and since it predicts performance in the corresponding 2-AFC task.

4 Motion detection by MT neurons and psychophysical performance

We now present a series of electrophysiological and psychophysical experiments aimed at understanding better the relation between the activity of single neurons and perception in the context of a motion detection task. These experiments are similar in spirit to the ones described in sect. 2. An important difference is that both electrophysiological recordings and behavioral experiments were performed simultaneously in trained awake behaving monkeys, thus allowing a direct comparison of the single neuron responses with behavior.

Experimental configuration. Monkeys were trained to perform a rating task in which dots moved within a circular window on a video screen (Fig. 6). A fraction of the dots was updated from frame to frame in such a way as to move coherently in a specified direction while the remaining dots were updated randomly. The stimulus was presented for a period of 2 sec and the animal was to report the direction of motion of the dots by making an eye movement towards one of two lights at the end of the trial. By changing the level of coherent motion, the difficulty of the task could be varied from easy (100 % coherence) to difficult (close to 0 % coherence, i.e. random motion of the dots in any possible direction).

In most experiments, the activity of neurons in the middle temporal area (MT) was recorded simultaneously during the task. MT neurons receive inputs from V1 and most of them (∼ 90 %) are directionally selective. Their responses are thought to be well described by the motion-energy model described earlier. The receptive fields of MT neurons are typically considerably larger than those of V1 neurons suggesting a convergence of information from V1 cells with different receptive fields.

During the recordings, the direction of preferred motion of the recorded cell was first determined and the stimulus was displayed in a circular region optimally covering the cell’s receptive field. The direction of dot motion was matched to the preferred or anti-preferred direction of the cell, to maximize the likelihood that the recorded cell contributed to the motion detection task.

Average neuronal and behavioral performance. Figure 7a illustrates the responses of an MT neuron during the presentation of stimuli of increasing coherence. Responses are typically variable both for preferred and anti-preferred motion direction. The number of spikes per trial is well fitted by a Gaussian distribution, a situation similar to that of
example 2 in sect. 3. As the coherence of dot motion is increased, the two distributions of spikes became better separated, thus conveying more information on the presence of preferred vs. anti-preferred motion stimuli. The animal’s performance in detecting preferred motion direction during the task is illustrated in Fig. 7b (open dots and dashed line). In this experiment, correct performance reached an 82 % threshold for 6.1 % motion coherence.

To compare the neuron performance with the observer’s performance, the authors computed the area under the ROC curve obtained from the spike number distributions for preferred vs. anti-preferred motion at each coherence level. This data is plotted as filled dots connected by a solid line in Fig. 7b. The neuron’s performance reaches 82 % performance at a coherence level of 4.4 % and is thus performing the task better than the monkey observer! This result is similar to that reported in sect. 2 for retinal ganglion cells. Typically, the neuronal performance determined by this method was within a factor 2 of the psychophysical performance for 76 % of the neurons. As we have seen in sect. 3, the area under the ROC curve is a measure of correct performance for a 2-AFC task rather than a rating task. Thus, the neurometric curve plotted in Fig. 7b should ideally be compared with the psychophysical performance in the corresponding 2-AFC task. It is however considerably more difficult to train animals to do 2-AFC tasks rather than rating tasks. Another interpretation is to assume that during a single rating trial, neuronal decision are based on two neurons with opposite preferred motion directions, but otherwise identical spike number distributions in response to the motion signals. If this were the case, listening to these two neurons and basing a decision on differences in their spike number would be sufficient to reproduce or exceed the animal’s performance.

**Lesion and microstimulation studies.** The visual cortex consists of a large number of areas besides V1 and MT and neurons sensitive to motion stimuli are found in many of these areas. Thus, it is entirely possible that the correlation between average neuronal performance and behavior described above is not due to a causal relation. An alternative possibility is that behavior is determined in another brain area and that MT neurons merely reflect the outcome of computations carried out in that area. Two methods can be used to make this alternative unlikely.

The first one consists in making a brain lesion restricted to area MT and measuring the behavioral performance of the animal before and after the lesion. Such an experiment is illustrated in Fig. 8. Fig. 8A shows the threshold for 82 % correct performance in the motion coherence task before and immediately after the lesion to MT for a range of dot motion speeds. The threshold is increased by about a factor 10 for a large range of speeds. Fig. 8B shows that this effect is specific: if the motion stimulus is presented in the opposite half of the visual field motion information will be represented in the MT area located on the opposite side of the brain. Since that area was not lesioned, one would expect unchanged performance if the effect of the lesion were specific to the lesioned area. This is indeed the case. This experiment shows that area MT is necessary to perform the psychophysical task: although decisions may not be taken in MT, they are based on information computed there.
Another experiment consists in electrically stimulating neurons in area MT while the monkey is performing the psychophysical task. In such an experiment, a cell is recorded in MT, its preferred motion direction is determined and the stimulus direction of motion is adjusted to match the preferred direction of the cell. During the task, a small electrical current is passed on a random subset of the trials. Such an electrical current is expected to excite neurons around the recording electrode. Since neighboring neurons will have similar response properties, this manipulation is expected to increase the activity in a population of cells with identical motion direction preference as the recorded cell. Fig. 9 A and B show two examples of such microstimulation experiments. In both cases the psychometric curve during microstimulation is shifted towards lower positive coherence levels. Thus, during microstimulation a higher fraction of correct responses is obtained even if the coherence is lower. This suggests that microstimulation can replace coherence as a factor contributing to increased performance. In Fig. 9A, the effect of microstimulation on psychophysical thresholds is equivalent to adding 7.7 % coherence to the dots while in B the effect was even higher (20.1 % coherence). This experiment suggests that altering the response properties of a small pool of cells (probably on the order of hundreds of cells) influences motion perception. Thus, the psychophysical decision are expected to be taken using the signals from this pool of neurons.

**Trial-by-trial correlation of neuronal responses and behavior.** The analysis presented in Fig. 7 compares the average performance across trials of single neurons with animal performance. It does not, however, address the question of whether single neuron responses covaried with behavioral responses on a trial-by-trial basis. In other words, at a fixed level of coherence such as 3.2 % in Fig. 8a, the animal is performing a sizable number of mistakes when the stimulus is moving in the preferred direction of the cell. If one breaks down the distribution of firing rates to that stimulus (stapled histogram bars) according to the animal’s responses, is there a correlation between responses and firing rate? Such an analysis is presented in Fig. 10. As may be seen from the figure, at each of the levels of coherence considered, the distribution of spikes for preferred direction decisions is slightly shifted towards the right as compared to the distribution of spikes for anti-preferred motion direction decisions. However, in the 6.4 % coherence case, the separation between the two distributions seems considerably smaller than that for 3.2 % coherence in Fig. 8a. In these three cases, the area under the ROC curve is given by 0.57, 0.75 and 0.78, respectively. Across neurons, the area under the ROC curve for such distributions is equal to 0.56 % and thus the correlation between single neuron responses and behavior on a trial by trial basis is weak. Thus, no single neuron appears to have a large impact on a single decision, in spite of the fact that many of them are on average as reliable as the animal. This suggests that psychophysical decisions are based on pooling information from hundreds of neurons, rather than relying on the neurons that are most reliable for the task.
Figure Legends

Figure 1. Left: Cumulative Poisson distributions for count thresholds from 1 to 9. Right: Psychometric curve for 3 observers fitted with the cumulative Poisson distributions on the left. Adapted from Hecht, Shlaer and Pirenne, J. Gen. Physiol. 25:819-840 (1942).

Figure 2. Left: Fit of the Barlow model to the HSP data ($\alpha = 0.13$, $x = 8.9$, $k_0 = 21$. Right: Performance of one observer under two conditions one is which the subject was encouraged to report only ”seen” flashes ($k_0 = 19$) and one in which the subject was encouraged to report ”possible or seen” flashes ($k_0 = 17$). False alarm rates were 0 and 1 %, respectively. Other parameters as on left. Adapted from Barlow, J. Opt. Soc. Am. 46:634-639 (1956).

Figure 3. Responses of a single retinal ganglion cell to 5 quanta (average) of light. A: PSTH, 10 ms bin width, 100 repetitions. B: Pulse count distribution in the presence (solid outline) and absence (dotted area) of the stimulus. C: ROC curve, i.e. probability of $c$ or more spike in presence of ”dark light” only and probability of $c$ or more spikes in the presence of the flash plus ”dark light”. Arabic numeral indicate threshold values for $c$. Roman numerals and crosses indicate values for an ideal detector assuming $\alpha = 0.18$ and $x = 6.5$ and $k_0 = 1 - 4$, respectively. Adapted from Barlow et al., Vision Res. Suppl. 3:87-101 (1971).

Figure 4. Two types of psychophysical tasks used to investigate the detection of weak signals. In the yes-no rating task, only a single instance of the stimulus is presented, either $s_0$ or $s_1$ and the subject is asked to choose the stimulus presented. During the feedback period, the subject is informed on his performance (for example a high tone is used to indicate a correct choice and a low tone for incorrect choices). In the 2-AFC task, both $s_0$ and $s_1$ are presented but their order of presentation is randomized. The subject is asked in which of the two intervals $s_1$ was presented.

Figure 5. Left: Receiver operating characteristic curves (ROC) for the ideal observer of two Gaussian distributions ($s_0$ and $s_1$) with equal variance and unequal means. The various curves correspond to different values of the normalized distance, $d$. Right: Performance of the ideal observer of two Poisson distributions with different means ($m_0$ and $m_1$ respectively). The circles and squares are labeled with the corresponding thresholds values ($k_{th}$) and linear interpolation between these values correspond to randomized tests built using the two nearest threshold values, as explained in example 1.

Figure 6. Schematics of the experimental configuration used in the Newsome experiments. Dots are presented in an aperture matched to the receptive field of the recorded MT neuron (bottom). The fraction of dots moving coherently in one direction is varied from zero coherence (top left, dots move in all possible directions) to 100 % coherence.
The stimulus is presented for 2 seconds and the subject is asked to answer the trial by making an eye movement towards one of 2 LEDs on the visual screen. The animal is required to fixate the fixation point throughout stimulus presentation. Adapted from Britten et al. J. Neurosci. 12:4745-4765, 1992.

**Figure 7.** a: Distribution of the number of spikes of a MT neuron for three different correlation levels (60 trials per correlation level; hatched bars: preferred direction, black bars anti-preferred direction). b: Neurometric curve (solid points and line) derived from a. The significance of this curve is explained in the text. The open dots and dashed line indicate the psychometric curve of the subject during the same trials. In this case, 82 % performance is reached for a 4.4 % coherence in the neuron and 6.6 % coherence for the animal. Adapted from Newsome et al., Nature 341:52-54.

**Figure 8.** Effect of lesions of area MT on the threshold coherence needed to detect reliably the motion stimulus (defined as 82 % correct performance). The left panel illustrates thresholds before and immediately after the lesion. The right side are the thresholds when the visual stimuli are presented in the opposite side of the visual field, which is processed by the MT area on the opposite side of the brain. Adapted from Newsome and Pare, J. Neurosci. 8:2201-2211, 1988.

**Figure 9.** Effect of electrical microstimulation in MT on the performance of a rhesus monkey in the direction discrimination task. Correlations on the abscissa are positive if they correspond to the preferred direction of the neuron and negative otherwise. The filled dot correspond to decisions taken during microstimulation trials and the open dots to decisions taken during trials without stimulation. In A microstimulation leads to a shift of the psychometric function leftwards corresponding to 7.7 % correlated dots. In B, the shift is much larger, 20.1 % correlated dots. Adapted from Salzman et al., Nature 346:174, 1990.

**Figure 10.** Distribution of spikes of an MT neuron at 3 different coherence levels as a function of the psychophysical responses. Correct responses correspond to the upward stapled spike distributions and incorrect responses to the downward spike distributions. Adapted from Britten et al. Vis. Neurosci. 13:87-100, 1996.
Yes-no rating task

---
stimulus presentation  response  feedback

2-AFC task

---
stimulus presentation  stimulus presentation  response  feedback

Figure 4

Figure 5

Figure 6