Old news and new news about single-photon sensitivity

Phil Nelson
Univ Pennsylvania
www.physics.upenn.edu/PtN

Image of chick retina by Andy Fischer, Ohio State Wexner Medical Center.
Whole-animal experiments show that our eyes are super-sensitive, and single-cell recordings showed that in fact, our photoreceptors can report single photon capture events, but as a physicist, I thought it important to ask whether those results could be quantitatively reconciled into a single model, even a laughably simplistic one.
Plan

1. Indoctrination
2. Light
3. Vision: psychophysics
4. Vision: single-cell
5. Synthesis
Eyes are an ancient invention: Here is Trilobite, half a billion years old.

That design was successful: Here is a modern aphid.

But if you can afford to carry more weight around, here is a better design:
The human eye has a lens-based focusing system, in some ways like a camera. It seems to make sense in terms of Snell’s Law, a consequence of the wave theory of light.

The finite size of our pupils limits the resolution we can get at the retina, due to diffraction. There’s no point having a pixel size smaller than this resolution limit, and remarkably, our photoreceptor cells really are about this size.

Looks like the wave theory of light explains everything.

Right from Polyak, *The vertebrate visual system* (1957)
Uh-oh

But what happens next? What happens in those photoreceptor cells that translates light into nerve impulses?

We can detect very dim light with a photomultiplier tube, or a photodiode. Either way, light causes discrete clicks in the detector. Dimmer light gives equally big clicks, just less frequent.

These are not uniformly spaced blips. Instead the clicks are as random as possible -- they are a “Poisson process.” Something about light is intrinsically random.

Moreover, when we shine dim light on several photodetectors, they never respond in unison: Each click comes from just one detector, even if the beam of light is spread out to cover them all.

Both digital and film cameras also expose one pixel at a time, at random:
Even classic diffraction effects turned out to be particulate in character.

Like the pattern of illumination in a focused image, a diffraction pattern turns out to be a probability density function for where the next pointlike blip will arrive.
What we’re facing

Not only is the emission of light lumpy and random: Also, when light cruises by a single atom or molecule, its capture (absorption) is also random – a “Bernoullii trial.” And typically the probability of capture is very low.

And how can an incoming photon “know” whether it “should” be absorbed by any particular molecule? It can’t – it must just make a random choice.

So if we find that our own eyes are extremely sensitive, then we must expect that randomness will play a big role in dim-light vision. How can we do science under such intolerable conditions, with such intrinsic irreproducibility?
Flash of light

Photoreceptor cell packed with sensitive molecules

Identically prepared flash of light

Identical photoreceptor cell

And again
Plan

1. Indoctrination
2. Light
3. Vision: psychophysics
4. Vision: single-cell
5. Synthesis
“What’s all that theoretical stuff got to do with vision, a real biological process? Surely the inconceivably tiny energy in a single photon is irrelevant to a macroscopic organism like me?”

But on the contrary: If your competitor/threatening predator/prey has better night vision than you, that could be a problem. Survival is the mother of invention, big time. If there’s a fundamental limit to vision, you need to be there if at all possible. Big eyeballs can help, but we can’t all carry around basketball-sized eyes like Architeuthis. We need to detect every photon, or come as close to that ideal as we can. Lorentz, 1920: Are we anywhere near to that ideal?

Answer: Reasonably close. Already at that time, existing measurements suggested that it takes about 25-200 photons to make a reliably-seen flash.
Hecht and coauthors made a more careful measurement. They sidestepped the “unrepeatability” problem by measuring the probability for a human subject to see a flash, vs intensity. They found that, at the lowest detectable flash intensities, seeing follows a sigmoidal curve. What can we conclude?

Hecht, Shlaer, Pirenne, JGP (1942)
An amazing deduction

Hecht et al said, “If 150 photons to the eye is enough to see a flash fairly reliably... and half of those are lost before they get to the retina (as is measured)... and 4/5 of the reminder are absorbed by something other than rhodopsin... then around 15 absorbed photons is enough to get a reliable conscious impression.

But those 15 are spread over a spot served by several hundred rod photoreceptor cells... so it’s very improbable that any one cell caught more than one photon... and yet this is a stimulus that reliably evokes a response... so it must be the case that an individual rod cell can generate a signal upon absorbing a single photon.

That is, we are all single-molecule biophysicists. right down to the lowly dung beetle.
In response to a faint flash, each individual rod photoreceptor either makes a signal or does not, independently with a probability proportional to the flash strength. But by itself, this model cannot fit the probability-of-seeing data, no matter what value we assume for that constant of proportionality.

So Hecht et al. assumed further that some neural system only alerts the conscious brain whenever the number of such signals exceeds some “network threshold,” $t$. That is, the probability of seeing is a cumulative Poisson distribution.

Thus, their hypothesis has two fit parameters: The probability per incident photon, $Q$, that a photoreceptor will signal (“quantum catch”), and the network threshold $t$.

% data from hecht et al page 835 table V 4th column
p=[0,0,0.12,0.44,0.94,1]; nphoton=[23.5,37,59,93,149,239];lnphoton=log10(nphoton);
plot(lnphoton,p,'o');
xlabel('log10 nbar','FontSize',16);ylabel('P_{see}','FontSize',16);title('probability of seeing','FontSize',16)
figure
plot(lnphoton,p,'--');hold on
for mstar=1:2:12,
    q=0.12;
    for j=1:46,
        photons=20+5*j;
        mbar=q*photons;
        total=0;
        for i=mstar:50
            total=total+exp(-mbar)*(mbar^i)/factorial(i);end
        ptheory(j)=total;photontheory(j)=log10(photons);end
    plot(photontheory,ptheory,'r');end
end
Sometimes simple is too simple

One often hears the takehome from the Hecht experiment expressed as “Our remarkable eyes can detect a single photon.” Unfortunately that won’t do:

- The Hecht data on a previous slide shows that, for reliable seeing, we need to present about 150 photons to the eye – not 1. Why that number? We partially took that into account when we included losses from absorption...

- Still, 150 x 10% = a mean catch of 15 photons for reliable seeing (or a bit less for 60% reliable). Why not 1?

So yes, our eyes are remarkable, but we still need the precise statement of how remarkable. Only when we’ve got that can we start to discard some hypotheses about what’s going on and retain others.
A better experiment

Also there’s a methodological problem in the Hecht experiment: What exactly does it mean to instruct subjects, “Don’t make false positive reports”? Van der Velden did a similar experiment independently, and got different results.

Barlow** realized that the discrepancy could be due to the different level of false-positive reports tolerated in each experiment. He proposed, and Barbara Sakitt then performed, a better psychophysics experiment than Hecht’s: She measured a function of two variables, the probability of a subject assigning each subjective rating to the strength of a flash as a function of that flash’s nominal strength. That gives a richer dataset:

---

Next turn to the puzzles in the model, such as the high network threshold. Wouldn’t I get better vision – and its fitness payoff – if I were aware of every photon absorption? Barlow pointed out that Hecht’s model omitted something big: the role of spontaneous isomerizations, a trickle of fake photon signals, indistinguishable from the real thing, that get folded into each rod cell’s output at some fixed rate.

Barlow suggested that our low-level visual circuitry applies coincidence detection to avoid distracting the conscious mind with too many of these of false positives. He noted:

• The combined signal from both kinds of isomerization is again a Poisson process, but with mean rate given by a sum:
  ✴ One contribution is the flash strength times quantum catch $Q$, as before;
  ✴ The other contribution, $\mu_0$, is independent of flash strength.
Barlow then proposed that

- As previously, some neural process downstream of the photoreceptors applies a network threshold to the total number of photon-like rod cell signals that follow a flash. (If there are several allowed responses, as in Sakitt’s experiment, then each one has its own threshold.)
- The low-level hardware is characterized by fixed parameters (summarized by \( Q \) and \( \mu_0 \)). But the high-level firmware is flexible; it can set threshold(s) suitable for the assigned task.
Barlow’s model sounds great—we just have more fit parameters than before, no problem for fitting software. I got a nice fit to all of Sakitt’s data with this model: I obtained a good-looking 8-parameter fit to more than 8 data points:

So, do we pop the champagne now? No, we say the model is promising. Unfortunately, it’s not very falsifiable. Many different fits to the data are possible, with widely varying values of the parameters (Fred Rieke).
So, do we pop the champagne now? No, we say the model is promising. Unfortunately, it’s not very falsifiable. “Just one more fit parameter” is suddenly a lot: In fact, many different fits to the data are possible, with widely varying values of the parameters (Fred Rieke).

So there’s more to be done:

Is the fit value of quantum catch $Q$ reasonable?
* Is there some independent way to obtain the value of the quantum catch, without fitting to psychophysical data? That would certainly make it harder for our model to give good-seeming results while actually being wrong. That is, it is good science to overdetermine the model as much as possible.

Is the fit value of spontaneous false-positive rate $\mu_0$ reasonable?
* Is there some independent way to obtain the value of the spontaneous false-positive rate?

Are the fit values of the network thresholds reasonable?
* E.g. they have got to be small integers – not much freedom – quite possibly no reasonable values could be found that give a fit after we constrain the other parameters.

Is everything after the initial photon catch really deterministic?
* We’d better look into alternatives.
Hecht et al.’s historic experiment did imply that, remarkably, a single productive photon absorption can elicit a response from a rod cell, without fancy electronics, decades earlier than they had any “right” to get this result.

But experiments of this type by themselves admit a range of different explanations. Other kinds of experiments were needed to break this degeneracy.
Plan

1. Indoctrination
2. Light
3. Vision: psychophysics
4. Vision: single-cell
5. Synthesis
An individual rod or cone cell’s response can be measured by gently aspirating its outer segment into a pipette electrode and stimulating it with 500 nm light (green streak).

Baylor’s experiments showed that although some flashes were missed altogether, others gave stereotyped responses. The experiments also two kinds of noise: signals in the dark that look exactly like real single-photon responses (“spontaneous isomerizations”) and a lower-amplitude rumble called “continuous dark noise” (due mostly to spontaneous activation of phosphodiesterase).

Data courtesy Greg Field (2008).
Baylor et al. found the probability of “seeing” curve for single rods. They knew the density of rhodopsin, its absorption cross-section, the rod cell dimensions. So from this data they obtained a meaningful (if partial) characterization of how good our visual apparatus is:

- A single productive absorption suffices to trigger a discrete signal.
- The false positive rate $\mu_0$, though not zero, is small.
- The value of the quantum catch, $Q$, can also be read out from such experiments.

It’s the same style of indirect reasoning as Hecht et al. used: Even though there was no way to say how many photons were in any given flash, still we see there’s no way to fit these data to a model in which individual rods have a response threshold of $>1$ photon.

We need a numerical value for the quantum catch when light is presented axially to the rod cell. But Baylor et al. measured the quantum catch when light is presented *sideways*.

Let \( c \) = concentration of rhodopsin pigments
and \( a_1 \) = cross-section. Optical density measurements give \( a_1c \approx 0.044\, \mu m^{-1} \).

Averaging the absorption probability over the variable thickness of a cylinder of diameter \( d_{rod}^m \) (macaque) gives

\[
P(\text{absorbed (sideways)}) = a_1c(\pi d_{rod}^m/4).
\]

The quantum yield for signaling is then

\[
\phi_{\text{sig}} = \frac{P(\text{rod signal} \mid \text{absorbed (sideways)})}{P(\text{absorbed (sideways))}}.
\]

The numerator of this expression is the sideways quantum catch, which is what Baylor measured. So we can convert that measurement into a value for the quantum yield:

\[
\phi_{\text{sig}} = Q_{\text{side}}/\frac{1}{4}a_1c\pi d_{rod}^m \approx 0.35.
\]

The interest of \( \phi_{\text{sig}} \) lies in the fact that its value should be *the same regardless of how light has been presented* to the rod cell. Once a photon has been absorbed by rhodopsin, its original direction no longer matters:

\[
P(\text{rod signal} \mid \text{absorbed (axial)}) = P(\text{rod signal} \mid \text{absorbed (sideways)}) = \phi_{\text{sig}}.
\]

The Beer-Lambert law then gives (\( L_{rod}^h \) = length of human rod cell)

\[
Q_{rod} = P(\text{rod signal (axial)}) = P(\text{rod signal} \mid \text{absorbed (axial)})P(\text{absorbed (axial)})
= \phi_{\text{sig}}(1 - \exp(-a_1cL_{rod}^h)) \approx 0.30.
\]

Finally we must convert from the quantum catch for a single rod cell, with light arriving axially and focused to just cover its cross-section, to the realistic case of the complete retina. For a 35-year-old subject stimulated with 507 nm light at 7 deg away from central vision: \( Q_{rod,\text{ret}} = [\text{ocular media}] \times [\text{tiling}] \times Q_{rod} \approx 0.076 \).
A more detailed prediction

Baylor et al. graphed the probability of any rod response as a function of the flash strength. But an earlier slide before that showed that their rod responses can be broken down into a series of discrete strengths, strongly suggesting that each bin reflects 0, 1, 2, ... productively absorbed photons during a flash exposure.

That is, the full Poisson distribution of counts, not just the cumulative distribution above 0, is measurable and should be compared to experiment.
Fast-forward to 21st century

Recently, high-tech experiments confirmed directly that rod cells impose no threshold:

Zero incoming photons: pure continuous dark noise

One incoming photon: some additional population of events

Data from Phan, Krichevskiy, et al *PRL* 2014.
How can a cell monitor a hundred million light-sensing molecules, signal when *any one* catches a photon, and not make too many false-positive signals? The answer’s a long story, but my point is that nobody would have dreamed to *ask* the question – it’s so crazy a claim – were it not for the pioneering experiments that said, “crazy or not, it’s true.”

The first synapse

Let’s return to the higher level of signals and what we do with them.

Baylor et al also had a prescient insight: The very first synapse must discard some genuine photon signals. That’s a second great example of being way ahead of your time by using indirect reasoning.

The first synapse imposes a threshold

Baylor et al. argued that because so many rod cells’ signals are pooled, even a few extra false positive signals would be bad for our vision. So the first synapse sets a breakpoint to filter out continuous dark noise, cleaning up the signals fed to the next cells in the network (rod bipolar cells):

Rob Smith et al argued that the breakpoint should not be set at the value shown, but rather should be higher, where it corresponds to *discarding half of the real signals*.

The mechanism of this thresholding is that the G-protein coupled receptors in each rod bipolar cell are saturated in the dark, and hence do not respond to small decreases in the release rate of glutamate from their rod cell.

Recordings from mouse rod cells (before first synapse):

Recordings from mouse rod bipolar cells (after first synapse):

Data courtesy Greg Field (2008).

The pioneering single-cell experiments established that:

- One productive photon absorption is enough for a rod cell to signal.

The experiments also measured:

- The quantum catch for a rod cell illuminated sideways. But we saw how to convert this to the corresponding value $Q$ for the case of interest (axial illumination).

- The mean rate of spontaneous signaling $\mu_0$. Multiply the measured single-cell rate by the number of rods in a summation region and the duration of an integration time to see how many fake, photon-like signals could be confused with a real flash.

- The rumble of continuous dark noise, which makes it necessary to impose a breakpoint at the first synapse, even though it means discarding a lot of real signals. Thus the number of rod signals that cross the first synapse, and the number of spontaneous isomerization signals, are both cut by about 50%, while continuous dark-noise signals are eliminated altogether.

It’s time to ask whether Barlow’s model (suitably amended to account for the breakpoint) can explain Sakitt’s data, if constants that were initially fit parameters are frozen to their measured values.

Somehow this hadn’t been done previously.
Plan

1. Indoctrination
2. Light
3. Vision: psychophysics
4. Vision: single-cell
5. Synthesis
The modeling challenge

It may be crazy-ambitious to attempt to model human psychophysics, corrupted as it is by all the complexity of the conscious and unconscious mind, with Barlow’s ultra-simplified physical model – but let’s try, after first extending in a simple way to account for continuous dark noise.

Barlow’s model involves two modules:

- Poisson processes thinned by various losses
- Assume no random losses, just pool signals and apply network thresholds.

Barlow model on the tightrope

- Single-cell physiology sets a *lower* bound on the visual system’s loss and randomness.
- The incredibly good dim-light vision of an entire animal sets an *upper* bound.

*Is there any room between these bounds, and if so, how much?* Even if our model of the intermediate processing is naive, finding one model that works would establish a baseline for others.
From cartoon to model: outline

- In any time window, the signals crossing the first synapse are Poisson distributed, with mean related to the flash strength, reduced by absorption by ocular media, increased by spontaneous isomerization, and reduced further by discrimination at the first synapse. Some loss and randomness is inevitable here. The parameters describing this “transduction module” are fixed.

- Some neural circuitry can count how many signals were generated anywhere in a “summation region” that includes the spot, in an “integration time” of around 200ms.

- Some neural circuitry applies a network threshold (“coincidence detector”) to that total to decide if a “real” flash took place. “Don’t bother the conscious mind unless at least $t$ photon signals were generated in one integration time, in one summation region.” At first we’ll suppose this is a deterministic step. However, its parameter $t$ is programmable according to the requirements of the task. There can even be multiple $t$’s.
Some details

- Each flash sends in photons in a Poisson process with mean rate related to light intensity. *We know that number.* (It’s purely physics.)
- Each rod cell in the illuminated spot has a fixed, small, probability of productively absorbing each incoming photon. *We got that number (“quantum catch”) from Baylor and some geometry.*
- Each rod cell reliably initiates a “signal” in response to even one productive absorption. *Now confirmed.*
- Some neural circuitry can “count” how many signals were generated anywhere in a “summation region” that includes the spot, in an “integration time” of around 200ms. *We have quantitative estimates of the summation region size* and the integration time*.
- But the light-sensitive molecules in rod cells can also spontaneously convert to their modified form, generating a dribble of false positive signals. *We get that rate from Baylor.*
- The first synapse discards about half of all isomerization signals, both true and false ones. *That fraction has been estimated*.
- All that remains to be fit, then, are the network thresholds $t_i$ for each of the rankings in Sakitt’s experiment. These must form an increasing sequence of small integers. *And there are many more data points than t’s, so the model is falsifiable.*

* = not very precise estimates, however
• We choose some values for the unknown fitting parameters (the integer thresholds for each verbal response).

• We want a likelihood function that will tell us which set of values is the winner.

• That’s the probability that the experimental data that actually were observed, would have been observed in a world with those values of the fit parameters.

• For each trial we know the true flash strength, and hence by earlier derivations we know the mean of the Poisson distribution of rod signals passing the first synapse.

• We sum that Poisson distribution from 0 up to \((t_1-1)\) to get the probability that the subject reports a rating of 0. We sum it from \(t_1\) up to \((t_2-1)\) to get the probability that the subject reports a rating of 1, and so on.

• We compute the log-likelihood of the entire dataset, then repeat for other values of the thresholds.

• The we find the winning values and see whether they indeed give a “good” fit.
For this subject, the best fit has first threshold equal to *just two* signals crossing the first synapse. (Other subjects could also be fit, in every case with first threshold 2 or 3.) (Sakitt’s old data give similar fits.)

Two is pretty close to the absolute limit (that is, 1). This subject found it worth her while (sufficiently informative) to set her first threshold to that low value.

But for the highest rating (r=4, or “certain”), the threshold was *nine* rod signals, close to Hecht’s value of 7.

Moreover, no fit is possible if we change the model to add significant loss, or additional noise, after the first synapse. Everything beyond that point must be nearly perfect.

Understanding how that is possible is the next challenge/opportunity.

It did prove possible to reconcile psychophysics with single-cell experiments in the context of a very simple model.

It was not obvious from the start that this would even be possible. For example, it is not obvious a priori that Barlow’s separation into two “modules” is right.

Although today Barlow’s picture of the “decision module” looks laughably oversimplified, we can regard this project as characterizing it in a way that any future, realistic model must explain.

That is, we found a way to take whole-organism (psychophysical) data and disentangle from it the first stage of processing, by itself fairly well understood now, so that we can focus on the next stages.

Technical details make this an interesting and challenging, but still doable, project. The best fitting model is the one that maximizes a likelihood function defined by the rather limited experimental dataset.
Plan

1. Indoctrination
2. Light
3. Vision: psychophysics
4. Vision: single-cell
5. Synthesis
6. Wrap
What is a physical model, anyway?

Don’t want to get all philosophical on you. I say, *It’s a Tetrahedron*. Today I applied this approach to dim-light vision, but it’s useful to think of any modeling challenge in this way:

```
for j=1:46,
    photons=20+5*j;
    mbar=q*photons;
    total=0;
    for i=mstar:50
        total=total+exp(-mbar)*(mbar^i)/factorial(i);
    end
```

More specifically, physical models are a category in which the early steps are rooted in the known behavior of simpler, nonliving systems; a kind of modularity is used (and justified); and falsifiable predictions are made concerning the behavior as physical parameters are changed.

“Yadda, yadda... photons, yadda,... cumulative Poisson distribution...”

Thanks

This material is the subject of a new textbook (Princeton Univ. Press, 2017) (www.physics.upenn.edu/biophys/PtN). Things discussed today were also published in Physical Biology (2016).

Also see:


*A student’s guide to MATLAB for physical modeling* by Tom Dodson and PN (free at www.physics.upenn.edu/biophys/PMLS).

*A student’s guide to Python for physical modeling* by Jesse Kinder and PN (Princeton Univ. Press, 2015).