Neuron, Volume *104*

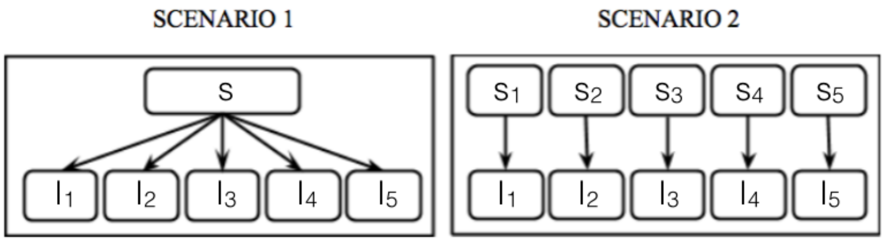
Supplemental Information

Bayesian Decision Models: A Primer

Wei Ji Ma

A Solutions to exercises

A.1 Case 1: Unequal likelihoods and Gestalt laws

a. 

b. $\frac{1}{2}$

c. $\frac{1}{32}$

d. The likelihoods do not add up to 1, nor should they. How probable the observations are under one scenario about the world is independent of how probable those same observations are under another scenario about the world.

e. The phrase is misleading because it suggests that there is only one. There are multiple different likelihoods for the same set of observations, one for each hypothetisized state of the world.

f. $p(H_1) = \frac{2}{3}$, $p(H_2) = \frac{1}{3}$

g. $\frac{1}{3}$

h. $\frac{1}{96}$

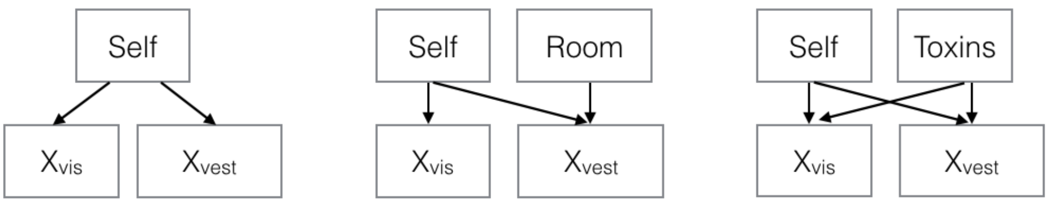
i. No.

j. $P(H_1|I_1, \ldots, I_5) = 0.97$, $P(H_2|I_1, \ldots, I_5) = 0.03$.

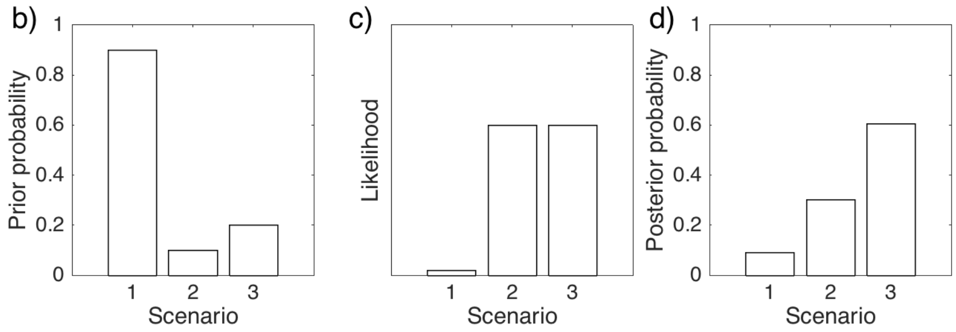
k. Since H_1 has the highest posterior probability, we are expected to perceive H_1– i.e., we perceive the group of dots as being part of the same object. This is consistent with the Gestalt law of common fate.

l. The Bayesian account produces predictions consistent with the Gestalt law but it goes beyond the "descriptive" law by providing an account that might be considered both normative and perhaps explanatory– it tells us what the optimal observer should perceive, beyond merely describing what people do perceive. Paired with certain evolutionary premises, the Bayesian account might provide an *explanation* of the law of common fate.

A.2 Case 2: Competing likelihoods and priors in motion sickness



a.



e. E.g. priors = $(0.9, 0.1, 0.2)$, likelihoods = $(0.01, 0.3, 0.3)$, posteriors = $(0.09, 0.3, 0.61)$.
f. Our evolutionary priors tell us stationary rooms are very probable, moving rooms are highly improbable, and the ingestion of a toxin is somewhat rare. The likelihood information we receive as a result of there being a visual-vestibular mismatch suggests that the stationary room hypothesis is improbable, and each of the other two hypotheses (which very likely lead to sensory mismatches) are probable. The posteriors are found by multiplying the prior and likelihood for each hypothesis. Scenario 3, the hypothesis that you are hallucinating because you've ingested a toxin, yields the highest posterior probability. If your body believes the MAP hypothesis that a toxin was ingested, it might trigger vomiting as a natural defensive response.

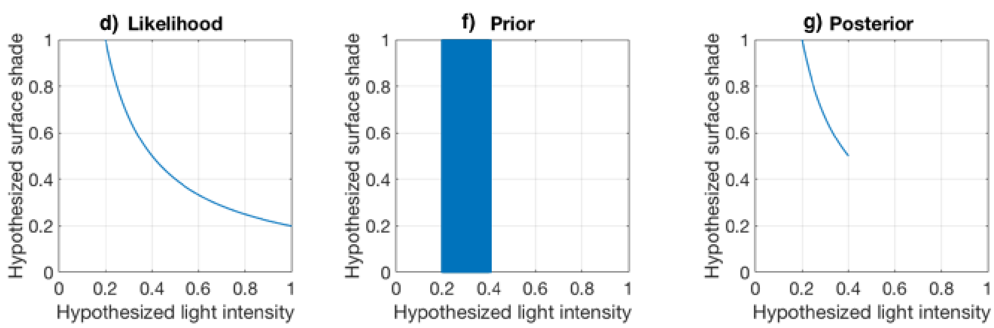
A.3 Case 3: Ambiguity from a nuisance parameter: Surface shade perception

a. 0.2

b. 0.5

c. Intuitively, the same retinal intensity might be caused by a bright light source and a dark surface , or a dark light source and light surface. Retinal intensity alone provides ambiguous information about surface shade, since we do not know the contribution of light intensity.

d.



Each plot should be interpreted as a heat map (with white background = 0).

e. This graph shows the likelihoods of each hypothesized combination of light intensity and surface shade, when the observed retinal intensity is 0.2. World states which lie off of this curve cannot produce a retinal intensity of 0.2. World states on this curve definitely produce a retinal intensity of 0.2.

f. To adjudicate between all possible points along this curve, the visual system might take prior over light intensity into account. For instance, if it is dark outside, there would be high prior probabilities for low light intensities. Combining with the likelihood, the posterior might be highest for the hypothesis that the surface of the object is white.

g. See figure.

h. See figure.

i. We would perceive the surface as having a shade somewhere between 0.5 and 1.

A.4 Case 4: Inference under measurement noise in sound localization

a. The proportionality sign is appropriate because s-independent factors are irrelevant when performing the final normalizing step needed to obtain the posterior.

b. The following piece of Matlab code approximately reproduces the figure:

```
clear, close all;
svec = -10:0.01:10;
mu = 0;
sig_s = 2;
sig = 1;
x_trial = 3.2;
prior = normpdf(svec, mu, sig_s);
prior = prior / sum(prior);
likelihood = normpdf(svec, x_trial, sig);
likelihood = likelihood / sum(likelihood);
protoposterior = prior .* likelihood;
posterior = protoposterior / sum(protoposterior);
figure; hold on
plot(svec,prior,'r')
plot(svec,likelihood,'b')
plot(svec,posterior,'k')
```

c. Posterior:

$$p(s|x_{\text{trial}}) \propto p(s)p(x_{\text{trial}}|s) \propto e^{-\frac{1}{2}\left(\frac{(s-\mu)^2}{\sigma_s^2}+\frac{(x_{\text{trial}}-s)^2}{\sigma^2}\right)}$$

We use the notation $J_s = \frac{1}{\sigma_s^2}$ and $J = \frac{1}{\sigma^2}$. Using the hint, we rewrite as

$$\begin{aligned} p(s|x_{\text{trial}}) &\propto e^{-\frac{1}{2}\left[(J_s+J)\left(s-\frac{\mu J_s + x_{\text{trial}} J}{J_s+J}\right)^2+\text{junk}\right]} \\ &\propto e^{-\frac{\left(s-\frac{\mu J_s+x_{\text{trial}}J}{J_s+J}\right)^2}{2\frac{1}{J_s+J}}} \end{aligned}$$

This is the equation for a Gaussian with mean and variance as given in the problem.

d. $\frac{\mu+x_{\text{trial}}}{2}$.

e. Should be straightforward.

f. Using the result of (e), $\sigma^2_{\text{posterior}} < \sigma^2$ because $\frac{\sigma_s^2}{\sigma^2+\sigma_s^2} < 1$. Analogously, $\sigma^2_{\text{posterior}} < \sigma_s^2$.

g. $\frac{\sigma^2}{2}$

h. As a rule for normal distributions, if $y \sim \mathcal{N}(\mu, \sigma^2)$, then $ay+b \sim \mathcal{N}(a\mu+b, a^2\sigma^2)$, e.g. scaling a Gaussian distribution by a factor of 2 would increase its variance by a factor of 4. The posterior mean estimate $\hat{s} = \mu_{\text{posterior}} = wx + (1-w)\mu$ involves scaling the measurement x by a constant w and shifting it by constant $(1-w)\mu$, where $w = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_s^2}}$. Therefore, the estimate distribution will have a variance which scales the variance of the measurement distribution by w^2:

$$\begin{aligned} p(x|s) &\sim N(x; s, \sigma^2) \\ p(\hat{s}|s) &\sim N(\hat{s}; ws + (1-w)\mu, w^2\sigma^2) \\ \text{Var}(\hat{s}|s) = w^2\sigma^2 &= \left(\frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_s^2}}\right)^2 \sigma^2 = \frac{\frac{1}{\sigma^2}}{(\frac{1}{\sigma^2}+\frac{1}{\sigma_s^2})^2} \end{aligned}$$

A.5 Case 5: Hierarchical inference in change point detection

a. The property that once the sequence $\mathbf{s}$ is given, $\mathbf{x}$ is independent of t_{change}. In other words, $p(\mathbf{x}|\mathbf{s}, t_{\text{change}}) = p(\mathbf{x}|\mathbf{s})$.

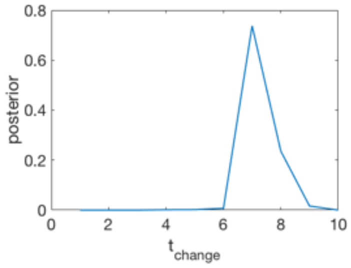
b. For a given t_{change}, only one sequence is possible.

c. Divide by the constant $\prod_{t=1}^{T} p(x_t|s_t = -1)$. This is a constant because it does not depend on t_{change}, and it can therefore be absorbed into the proportionality sign.

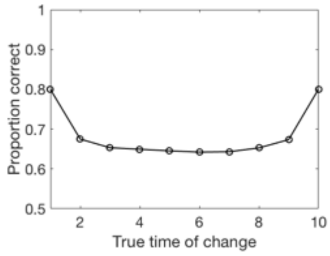
d.

$$L(t_{\text{change}}) \propto \prod_{t=t_{\text{change}}}^{T} \frac{p(x_t|s_t=1)}{p(x_t|s_t=-1)} = \prod_{t=t_{\text{change}}}^{T} e^{-\frac{(x_t-1)^2}{2\sigma^2}+\frac{(x_t+1)^2}{2\sigma^2}} = e^{\frac{2}{\sigma^2}\sum_{t=t_{\text{change}}}^{T} x_t}$$

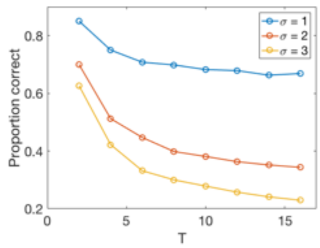
(f)



(i)



(k)



g. Normalizing does not change the most probable change point, since it is just the highest point on the graph.

j. 0.683.

l. The change point could have a higher probability of occurring at some times rather than others. There could be multiple change points instead of one. The stimulus could take more values than merely -1 or 1.