Models of Decision Making

Ramesh Srinivasan

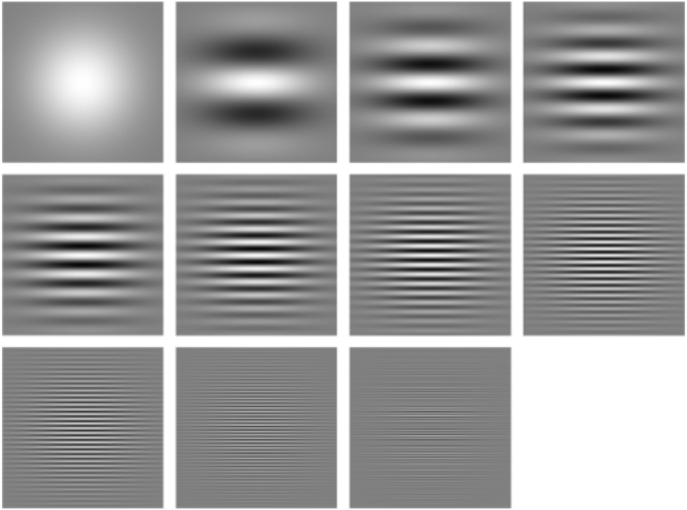
October 17, 2023

Decision Making Model

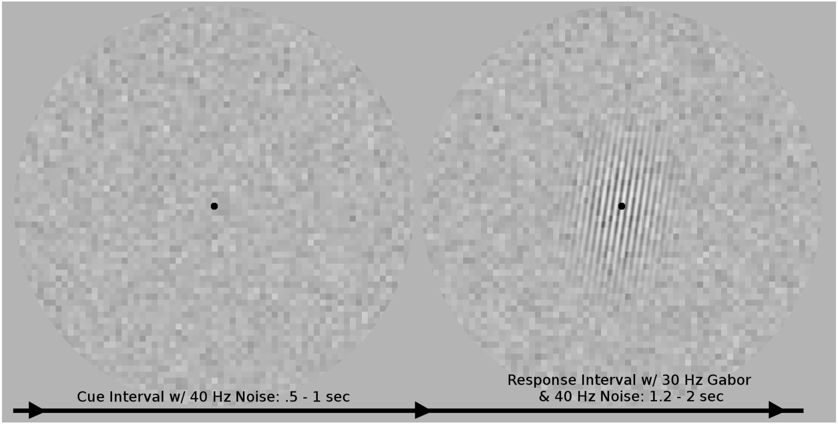
A verbal theory of response time and accuracy data in decision making

- There is a variable amount of time it takes for a stimulus to be processed by different stages of the sensory system and the objects in the environment to be perceived. This is often termed *figure-ground segregation*, and it is believed to require 150-250 ms.
- There is a variable amount of time it takes for the motor cortex to activate and send a signal down the spinal cord to the appropriate muscles and for the muscles to activate to provide a response. This is at least 50 ms, but depends on a number of factors and will vary between trials.
- In between perception and action there is a process which takes time, and potentially leads to errors, whereby the stimulus representation is analyzed, the stimulus is categorized and an action is selected. We label this collectively as decision making.

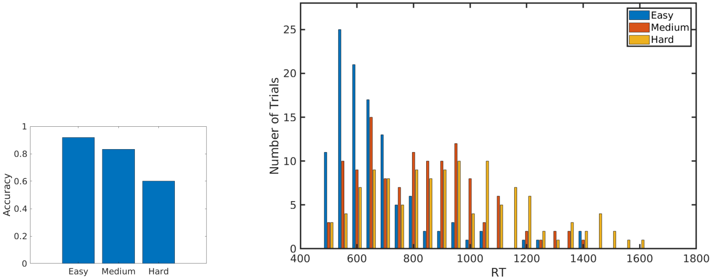
Spatial Frequency Discrimination



Example Stimulus Sequence



Example Data - Varying Task Difficulty

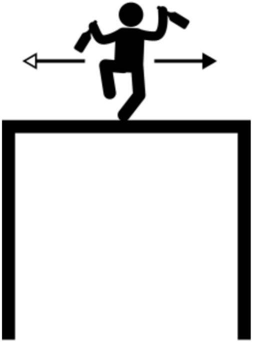


Easy, Medium, and Hard conditions correspond to the amount of *difference* in spatial frequency.
Thus the amount of discriminatory information available in the image decreaes as the task becomes more difficult.

Hypothesis about Two-Alternative Decision Process

- Assume a model that samples evidence in discrete time windows. The evidence obtained in each time window will favor one of the two alternatives (a "nudge").
- Each sampled number represents a nudge toward one decision or another.
- The magnitude of that nudge reflects how much information is acquired in that single sample.
- The nudge would then be added to the sum of previous nudges, moving a decision variable towards one or the other alternative. (An alternative view is that there are *two* decision variables involved in a *race*)
- A decision is made when the decision variable crosses a boundary (In a race or competing accumulator model, one of the two accumulators crosses the boundary).

Random Walk



- Random walk models originated in physics and chemistry (e.g., Brownian motion of a gas)
- They are used in psychology and neuroscience.
- They are also used in economics and finance.

Random Walk Simulation

Random walk model:

1. Suppose the random walk process starts at X_0 (This is an important variable that expresses *bias*!)
2. Suppose X_t is the step at time t. X_t can be any positive or negative number, randomly drawn from some distribution
3. The evidence, i.e. position of the walker, after 1 step is defined as: $Z_1 = X_0 + X_1$
4. The evidence, i.e. position of the walker, after 2 steps is defined as: $Z_2 = X_0 + X_1 + X_2$
5. More generally, after T steps

$$Z_T = \sum_{t=0}^{T} X_t$$

6. In most programming languages that expression is known as a cumulative sum and can be executed in a single step.

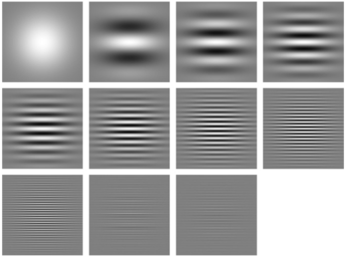
Drift, Bias, and Boundaries

- In our example of the drunken walker, there are two boundaries where he falls of the cliff. The "decision" the drunk takes is which side to fall off. The boundary represents a *criterion* for how much *evidence* is needed to select one decision.
- The sample on which the walker falls of the cliff is the *response time*
- We introduce the notion of *bias*, i.e., X_0. If the walk starts equidistant from the two boundaries we are equally likely to fall off one side or the other, but if we start closer to one edge, we will fall off that side more often.
- In real decision making there is some information in the stimulus. How can we represent the influence of information on the random walk decision model. The information available is represented as evidence pointing the walker to the left or right wall. One way we can represent this is for the *mean* step or nudge, labeled *drift rate* to be in the direction indicated by the stimulus.
- If the walker falls of the cliff in the direction indicated by the stimulus, the response was correct, but if the walker falls of the clip on the other side, the response was incorrect. Thus *accuracy* is incorporated in this model.

Ambiguity in the model - Drift, Diffusion, Criterion

The simulation carried out with Example 3.m varied the mean step and standard deviation proportionately, and was found to simply scale the random walk.
The average (mean of the normal distribution) step is labeled *drift rate*.
The variability in the walk (standard deviation of the normal distribution) is labeled *diffusion coefficient*
The scaling behavior of these parameters indicates that I can produce the same response time distribution by simply scaling, drift, diffusion and criterion, and there are an infinite combination of value that will produce the same response distribution.
To interpret the model one of these parameters has to be fixed, and by convention the diffusion coefficient is held fixed and criterion is varied.
This is because the focus of the use of this model is to study individual differences in RT, and its assumed that this comes from information processing ability (drift rate) and individual differences in strategy (criterion).
As an aside this makes me deeply unhappy, and I think it undersells the psychological meaning of the diffusion coefficient, which I believe can be related to attention.

Spatial Frequency Discrimination

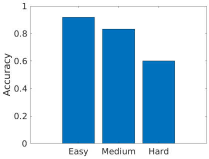


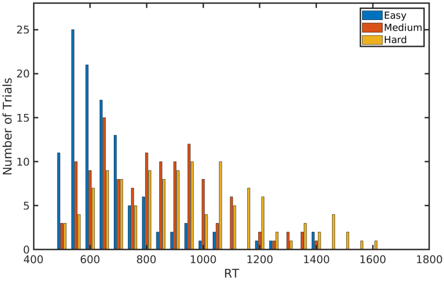
In the spatial frequency discrimination experiment, as the spatial frequency of the two Gabors to be discriminated become more similar we expect the drift rate to go down as the information available in the stimulus is reduced.
Increased task difficulty produced longer reaction times, which the model would suggest comes from lower drift rates.
You can think of drift rate as the average slope of a regression line fit to the random walks.

A verbal theory of response time and accuracy data in decision making

- There is a variable amount of time it takes for a stimulus to be processed by different stages of the sensory system and the objects in the environment to be perceived. This is often termed *figure-ground segregation*, and it is believed to require 150-250 ms.
- There is a variable amount of time it takes for the motor cortex to activate and send a signal down the spinal cord to the appropriate muscles and for the muscles to activate to provide a response. This is at least 50 ms, but depends on a number of factors and will vary between trials.
- **These two processes of perception and motor control add to the reaction time but are not directly part of the decision making**. They add a variable amount of time to the reaction time which our model must include.
- In between perception and action there is a process which takes time, and potentially leads to errors, whereby the stimulus representation is analyzed, the stimulus is categorized and an action is selected.

Example Data - Varying Task Difficulty



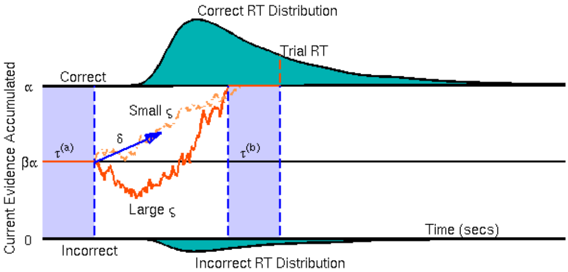


Easy, Medium, and Hard conditions correspond to the amount of *difference* in spatial frequency.
Thus the amount of information available in the image decreases as the task becomes more difficult, which we expect relates to drift rates.
Notice that in all 3 conditions, the fastest trials have similar RTs. **What does this likely mean?**

Non-decision time

To have a complete model of reaction time distributions, we need to include some perceptual and motor processing time.
The timing of these processes have clear correlates in EEG and single-unit recordings in animal models.
They are also clearly visible in the RT distributions as the fastest RTs observed.

Drift-Diffusion Model



$\tau = \tau^(a) + \tau^(b) = \tau_v + \tau_m$ Non-decision time

δ Drift Rate - Mean Evidence Accumulation Rate

ζ Diffusion Coefficient- Standard Deviation (variability) in Evidence Accumulation

α Boundary Separation or Criterion. Usually a positive parameter.

β Bias - Fraction between 0 and 1.

Ex-Gaussian Probability Density Function

The combination of the exponential and normal distribution is a convolution, resulting in the probability density function,

$$f(x|\mu,\sigma,\tau) = \frac{1}{\tau} e^{\frac{\mu-x}{\tau}+\frac{\sigma^2}{2\tau^2}} \Phi(\frac{x-\mu}{\sigma} - \frac{\sigma}{\tau})$$

where $\Phi(x)$ is the cumulative density function of the Normal distribution.

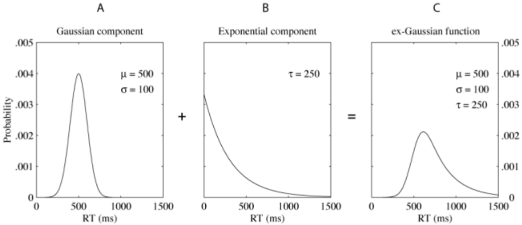
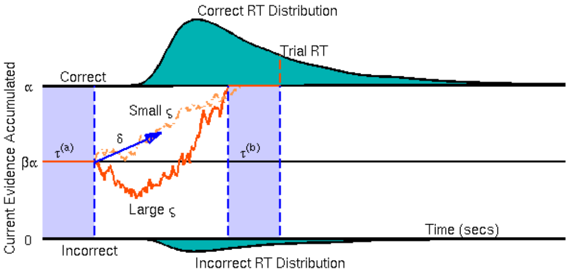


Figure 4. The ex-Gaussian probability function with parameters μ = 500, σ = 100, and τ = 250 (Panel c) resulting from the convolution a Gaussian probability function (Panel A) with an exponential function (Panel B).

Drift-Diffusion Model



$\tau = \tau^(a) + \tau^(b) = \tau_v + \tau_m$ Non-decision time

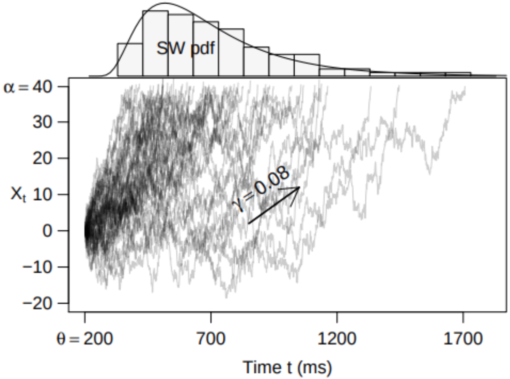
δ Drift Rate - Mean Evidence Accumulation Rate

ζ Diffusion Coefficient- Standard Deviation (variability) in Evidence Accumulation

α Boundary Separation or Criterion. Usually a positive parameter.

β Bias - Fraction between 0 and 1.

Diffusion Model - Single Boundary



In this simplified diffusion process, there are 3 parameters:
θ which is referred to as shift which is nondecision time.
α which is the boundary separation or caution.
γ which is drift rate or the rate of evidence accumulation.

Shifted Wald, or Shifted Inverse Gausian

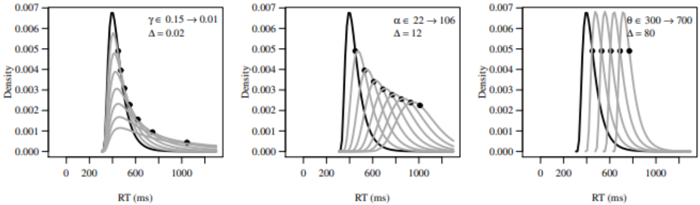


Figure 2. An illustration of how the SW distribution shape changes as one manipulates each parameter. Left to right, changes in γ, α, and θ, each in a direction that produces larger mean values. In the left plot, the black distribution starts with $\gamma = 0.15$ and each successive grey distribution is a reduction of 0.02 units, until γ reaches 0.01.

$$f(X|\gamma,\alpha,\theta) = \frac{\alpha}{\sqrt{2\pi(X-\theta)^3}} e^{-\frac{|\alpha-\gamma(X-\theta)|^2}{2(X-\theta)}}$$

Model Selection

Theoretical Basis

The first question we should ask in model selection is whether the model makes sense, and can be used to answer a scientific question of interest.

The second question we should ask is whether the complexity of the model is matched to the data. We have three main models of reaction time data presented here - the exGaussian model, the shifted Wald model, and the drift-Diffusion model.

The exGaussian model is the simplest model, but has the weakness of being mainly descriptive. That is the parameters of the model do not have a *cognitive process* that they can be related to.

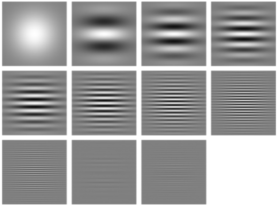
Experimental Evidence

The shifted Wald model is a richer model with a clear link to the process that generated the data. Each of the parameters has a meaning.

1. The shift parameter θ is related to the time outside of "decision making", i.e., evidence accumulation process. This is the time for sensory processing and motor response.
2. The boundary parameter α reflect the caution of the subject - how much information do they need to make a decision
3. The drift parameter γ reflect the rate of information accrual, or difficulty of the task.

The main "weakness" of the shifted Wald model is that it can only make use of **correct** trials and completely ignores the incorrect trials.
The full drift-diffusion model has more parameters to the shifted Wald model but maybe a better model as it incorporates the incorrect trials.
Which model to fit depends on the nature of the task and the data available. For example, in tasks with very high accuracy, the shifted Wald model may be a better model, as there are insufficient incorrect trials to fit the drift-diffusion model.

Framing Your Hypothesis Quantitatively in the Model



Lets consider the spatial frequency discrimination task whose data you have been looking at.
The data show that as the discrimination becomes more difficult (difference in spatial frequency becomes smaller) response time increases for accurate trials.
Using the single diffusion model (shifted-Wald) can we express some hypothesis about the data.

1. The experiment was designed to manipulate drift rate. As the spatial frequencies become closer there is less evidence to accumulate on each sample, so gamma should be smaller and reaction time increase.
2. However, there is always the possibility that boundary is varying with task difficulty as the subjects became more cautious when they find evidence accumulation is slower (task harder).

Shifted Wald, or Shifted Inverse Gausian

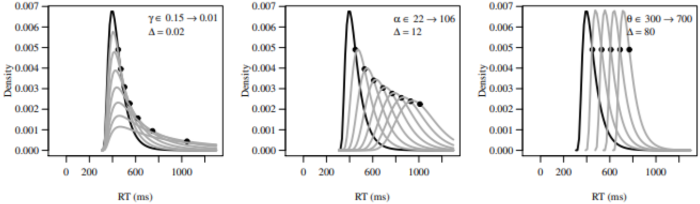


Figure 2. An illustration of how the SW distribution shape changes as one manipulates each parameter. Left to right, changes in γ, α, and θ, each in a direction that produces larger mean values. In the left plot, the black distribution starts with $\gamma = 0.15$ and each successive grey distribution is a reduction of 0.02 units, until γ reaches 0.01.

$$f(X|\gamma, \alpha, \theta) = \frac{\alpha}{\sqrt{2\pi(X-\theta)^3}} e^{-\frac{|\alpha-\gamma(X-\theta)|^2}{2(X-\theta)}}$$