- Why did you choose this course?
- Why did you choose the clothes you’re wearing?
- Why are you sitting where you are?
- Why are you reading this?

- Who or <u>what</u> made the decision???

Decision Making

Peggy Seriès, IANC
Informatics, University of Edinburgh, UK

pseries@inf.ed.ac.uk

CCN Lecture 7

- Why did you choose this course?
- Why did you choose the clothes you’re wearing?
- Why are you sitting where you are?
- Why are you reading this?

- Who or what made the decision???

Today’s lecture

1. **Decision theory**: what is the optimal way to make a (binary) decision? (optimal model)

2. **Behaviour**: How do human-beings make binary decisions? (phenomenological model of accuracy and reaction times)

3. **Neuroscience**: Can we relate the optimal model and phenomenological model with processes observed in neurons in the brain? (yes!)

1) Optimal model: statistical inference

- Decision making can be thought of as a form of statistical inference.
- Decide = select among competing hypotheses h_1, h_2 (and may be more). **Is $P(h_1|e) > P(h_2|e)$?**

- Elements of this decision process:

✳ evidence (e) = information we can collect in factor of h_1. Only useful when we know how likely it is to be true if the hypothesis is true, i.e. if we have conditional probabilities such as $P(e| h_1)$ = the likelihood

✳ priors ($P(h_1)$)= Probability that h_1 is correct before collecting any evidence = a bias (or prejudice)

✳ value (v) = subjective costs and benefits for each outcome.

Bayes' Theorem

- Bayes' theorem is a result in probability theory that relates conditional probabilities P(AIB) and P(BIA)
- Given the likelihood and the prior, we can compute the posterior.

Reverend Thomas Bayes, 1702- 1761

$$P(h_1|e) = \frac{P(e|h_1)P(h_1)}{P(e)}$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}}$$

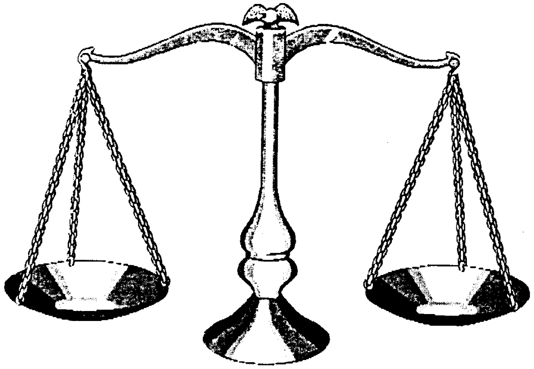
To decide, compare probabilities of each hypothesis

- Choose h_1 if:

$$P(h_1|e) = \frac{P(e|h_1)P(h_1)}{P(e)}$$

$$>$$

$$P(h_2|e) = \frac{P(e|h_2)P(h_2)}{P(e)}$$



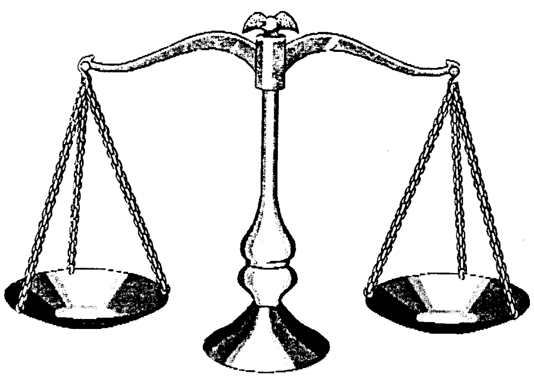
To decide, compare probabilities of each hypothesis

- Choose h_1 if:

$$P(h_1|e) = \frac{P(e|h_1)P(h_1)}{P(e)}$$

$$>$$

$$P(h_2|e) = \frac{P(e|h_2)P(h_2)}{P(e)}$$



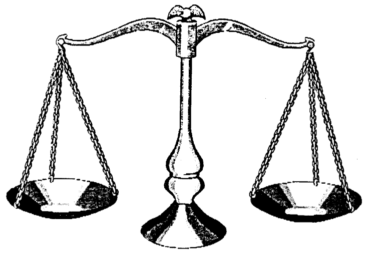
Likelihood ratio test

- Just re-organising the terms of this inequality: - **choose h_1 if:**

$$\frac{P(e|h_1)}{P(e|h_2)} > \frac{P(h_2)}{P(h_1)}$$

- This is the likelihood ratio (LR) test = optimal decision rule.
- If the prior probabilities are equal (0.5), choose h_1 if

$$LR = \frac{P(e|h_1)}{P(e|h_2)} > 1$$



Values (1)

• It might be that the costs and benefits associated with the various outcomes are very different.

• benefit of choosing h_1 =

value of choosing h_1 if h_1 is true (V_{11})

+ value of choosing h_1 if h_1 is wrong

(V_{12}) given the evidence.

• benefit of choosing h_2 =

value of choosing h_2 if h_2 is true (V_{22})

+ value of choosing h_2 if h_2 is wrong

(V_{21}) given the evidence.

run or not?

• So we now want to compare:

$$V_{11}P(h_1|e) + V_{12}P(h_2|e) \text{ with } V_{22}P(h_2|e) + V_{21}P(h_1|e)$$

Values (2)

- rewriting this gives the general (optimal) rule: **choose h_1** if :

$$\frac{P(e|h_1)}{P(e|h_2)} > \frac{(V_{22} - V_{21})P(h_2)}{(V_{11} - V_{12})P(h_1)}$$

- which has also the form of comparing the likelihood ratio with a threshold.

- Signal detection theory: LR (or any monotonic function of it - e.g. LOG) provides an optimal ‘decision variable’.

Sequential Analysis

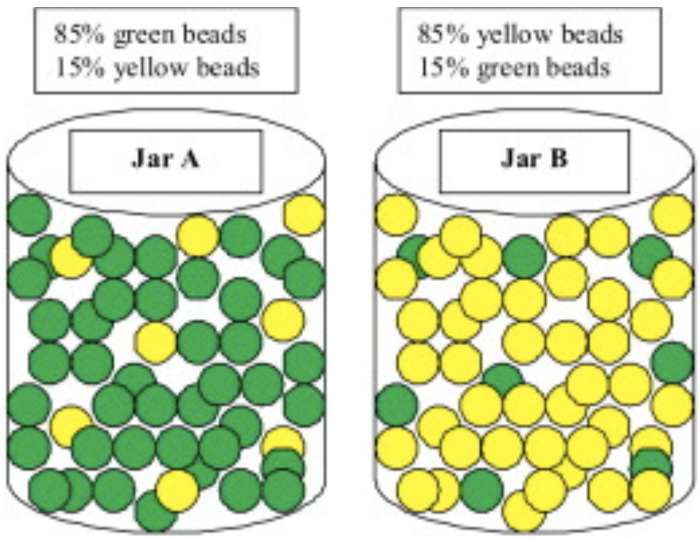
- This framework can be extended to the situation where we have multiple pieces of evidence e_1, e_2, ..e_n observed over time.
- Here we allow the decision variable to 'accumulate the evidence' in time:

$$\log LR_{12} \equiv \log \frac{P(e_1, e_2, \ldots, e_n | h_1)}{P(e_1, e_2, \ldots, e_n | h_2)}$$
$$= \sum_{i=1}^{n} \log \frac{P(e_i | h_1)}{P(e_i | h_2)}.$$

- When the DV > threshold A (which reflects priors and values), a decision is made towards h_1. If DV < B, choose h2.

- This is known as **Wald's sequential probability ratio test (optimal rule).**

Belief updating in the Beads (or Urn) Task



$P(G|\text{ jar A}) = 0.85$
$P(Y|\text{ jar A}) = 0.15$
$P(G|\text{ jar B}) = 0.15$
$P(Y|\text{ jar B}) = 0.85$

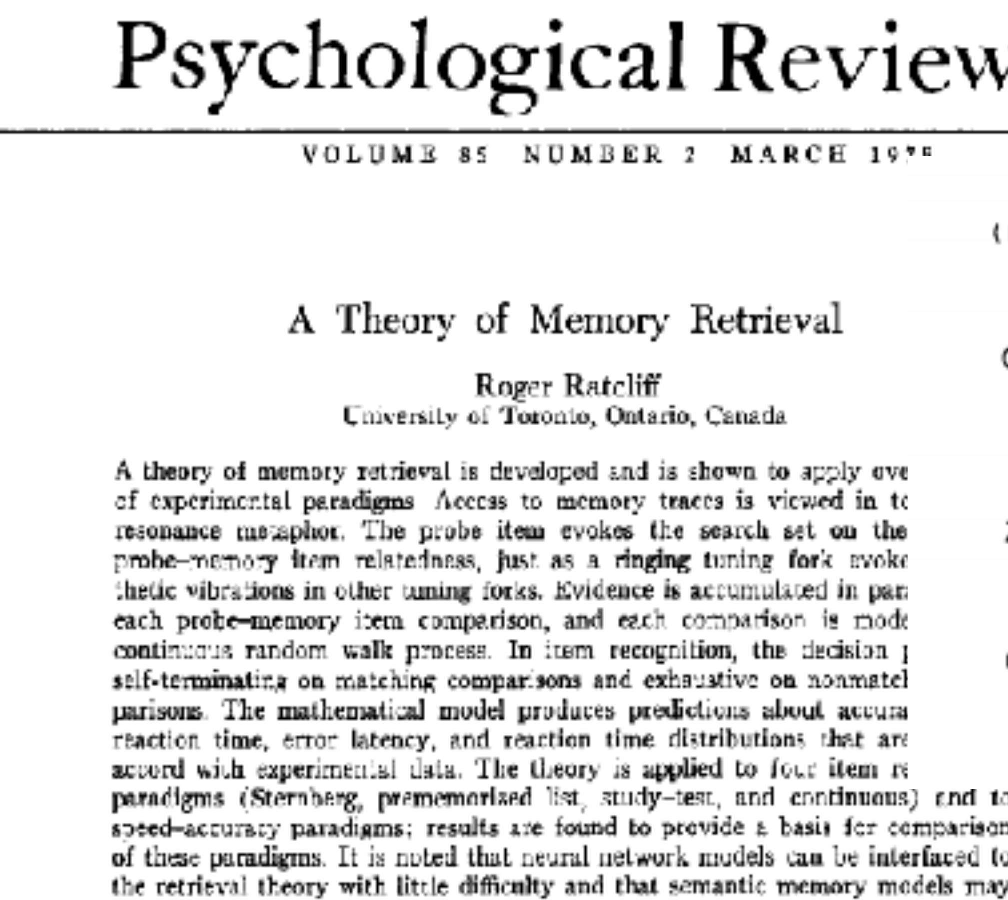
Which jar am I drawing from ?
When can you commit to a decision?

$$LogLR = log\frac{P(G|jarA)}{P(G|jarB)} + log\frac{P(G|jarA)}{P(G|jarB)} + ..$$

lnLR 1.72 +1.72 -1.72

2) How do humans make decisions?

• Sequential analysis is reminiscent of the random walk and race phenomenological models of decision making developed by psychologists to explain behavioural data.



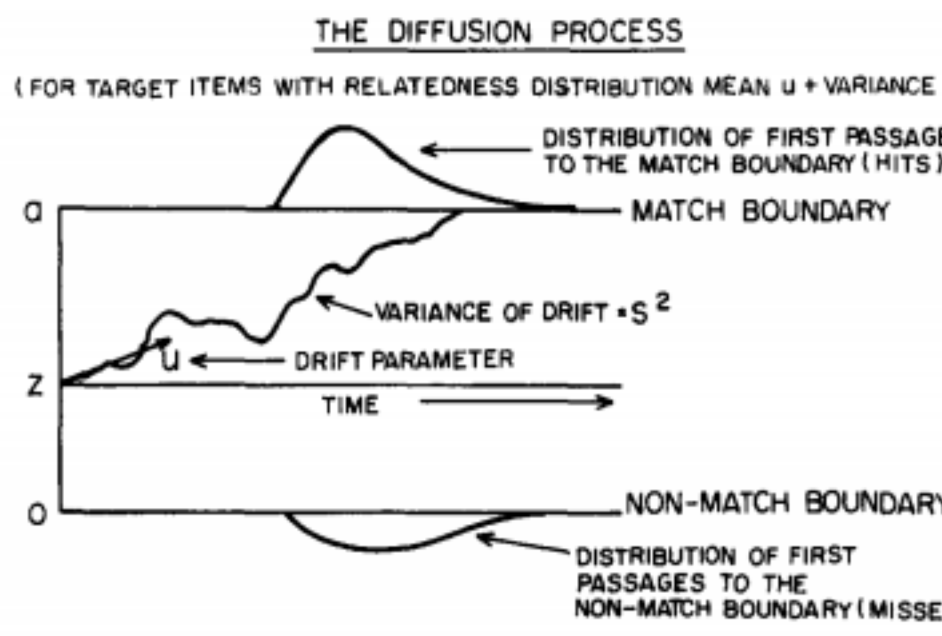
Psychological Review

VOLUME 85 NUMBER 2 MARCH 1978

A Theory of Memory Retrieval

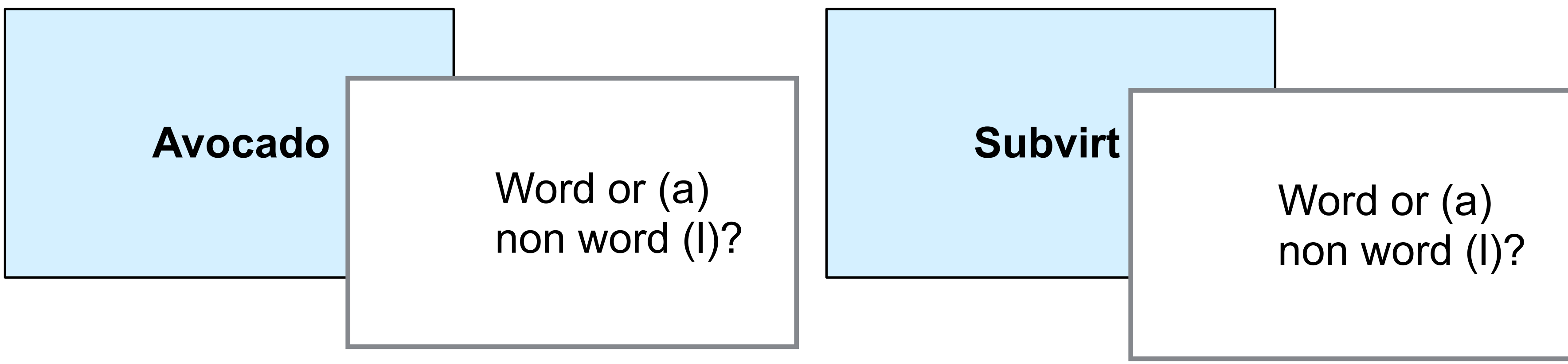
Roger Ratcliff
University of Toronto, Ontario, Canada

A theory of memory retrieval is developed and is shown to apply ove of experimental paradigms. Access to memory traces is viewed in te resonance metaphor. The probe item evokes the search set on the probe–memory item relatedness, just as a ringing tuning fork evoke thetic vibrations in other tuning forks. Evidence is accumulated in par each probe–memory item comparison, and each comparison is mode continuous random walk process. In item recognition, the decision p self-terminating on matching comparisons and exhaustive on nonmatcl parisons. The mathematical model produces predictions about accura reaction time, error latency, and reaction time distributions that are accord with experimental data. The theory is applied to four item re paradigms (Sternberg, prememorized list, study–test, and continuous) and to speed–accuracy paradigms; results are found to provide a basis for comparison of these paradigms. It is noted that neural network models can be interfaced to the retrieval theory with little difficulty and that semantic memory models may



How do humans make simple decisions?

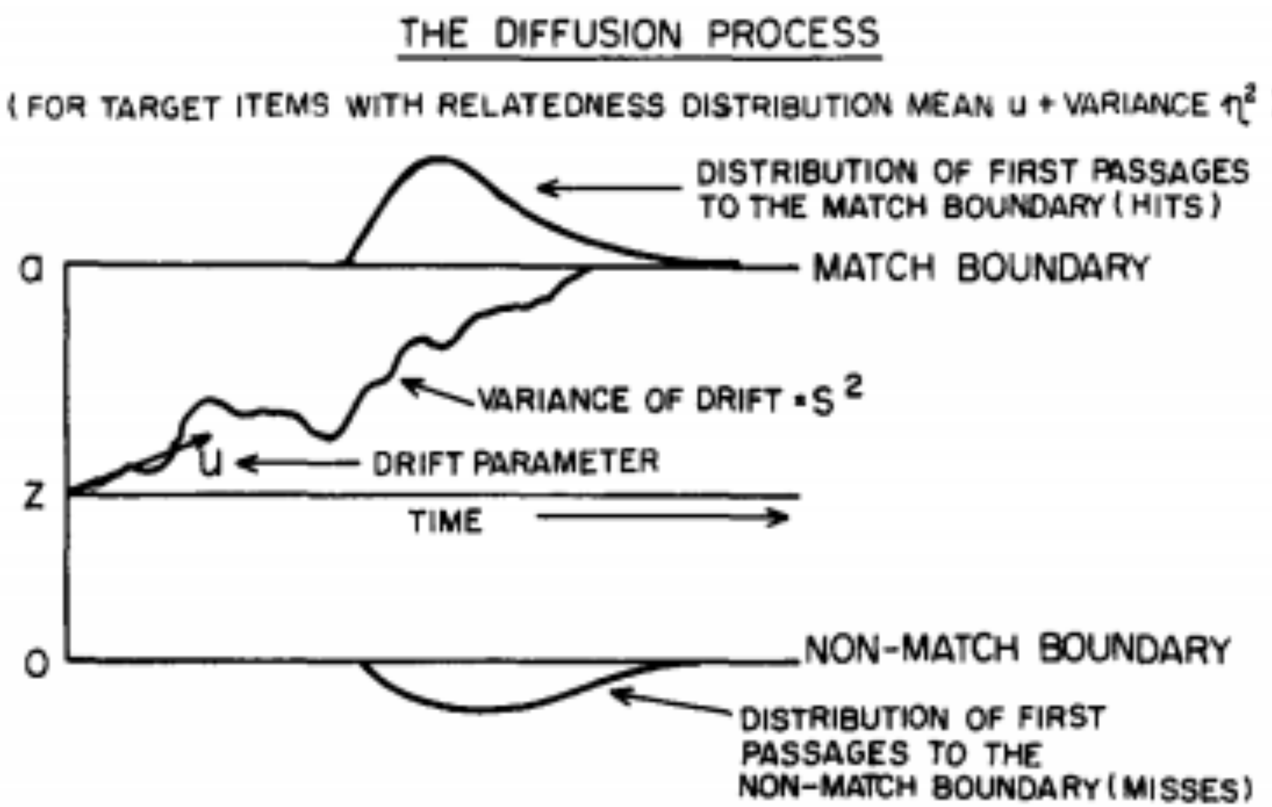
- Study of simple (single-stage), **fast** **(less than 2 seconds) binary decisions**
- Performance is described in terms of **reaction times** and **accuracy**.
- perceptual discrimination (are these 2 objects the same or different?), recognition memory (is this image new or was it presented before?), lexical decision (is this a word or a non word?
- How can we describe how fast and accurately people respond?



https://www.psytoolkit.org/experiment-library/experiment_ldt.html

How do humans make decisions?

• idea for the phenomenological model: decision can be viewed as resulting from the movement of a particle moving in between/two boundaries, pushed by the force of the evidence



Drift Diffusion Model of Decision Making

Several mathematical expressions exist for the DDM. A typical equation will be of the form of a **Wiener process** (one dimensional Brownian motion). The diffusion process *x(t)* evolves dynamically according to:

$$\frac{dx(t)}{dt} = v(t) + \sigma\eta(t)$$

- Where ***v*** is the **drift rate,** the quality of the information evidence. If the stimulus is easily classified, it will have a high rate of drift and approach the correct boundary quickly, leading to fast and accurate responses.
- **η** is a **white noise** term.
- σ^2 is the variance of the process.
- **starting point**, ***z***, to one of two **boundaries**, ***a**,* or ***0***.
- The two boundaries represent the two possible decisions. **Once the process *x(t)* reaches a boundary, the corresponding response is initiated.**

Drift Diffusion Model of Decision Making

Another form is the **Ornstein-Ulhenbeck process:**

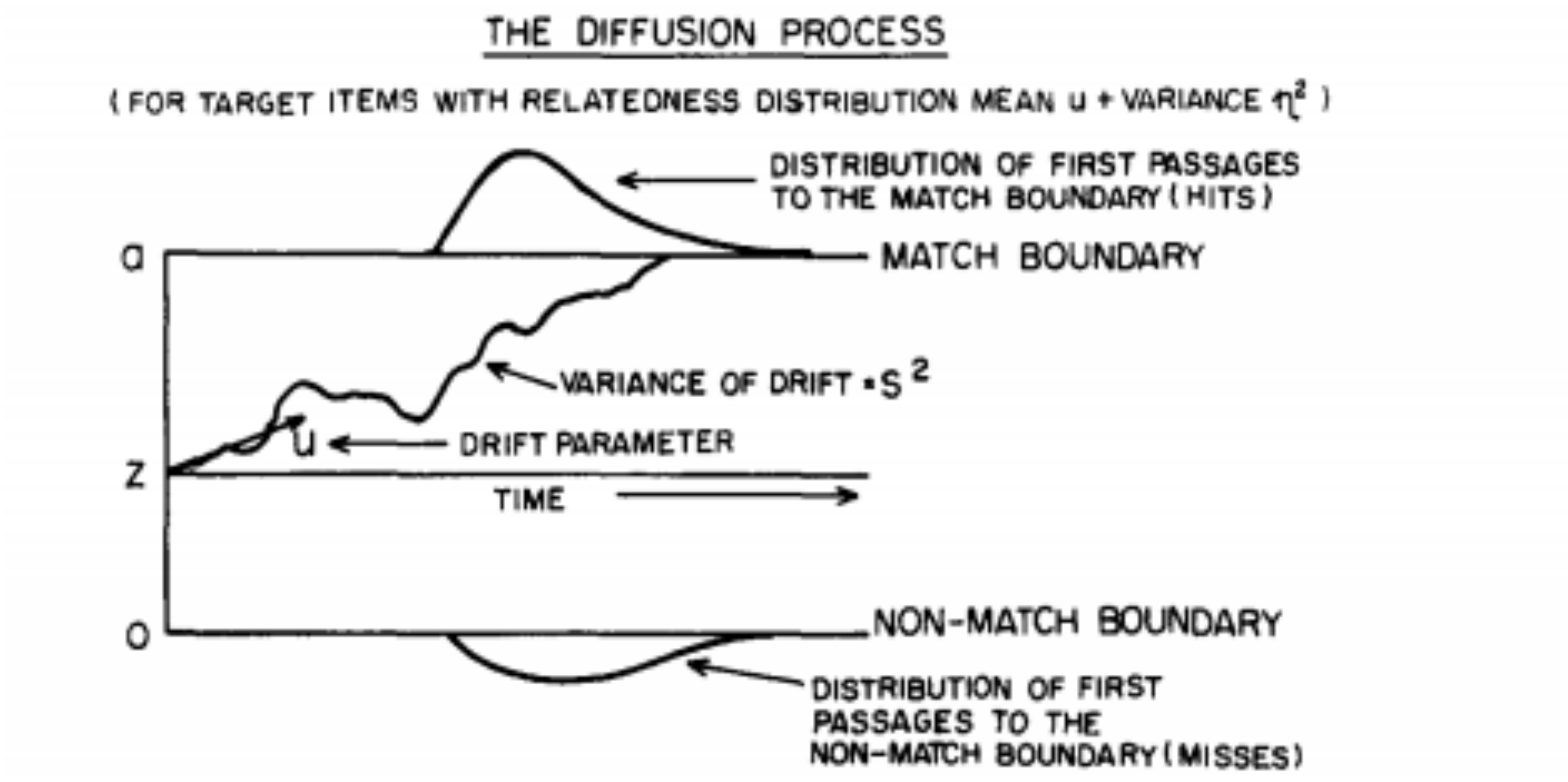
$$\frac{dx(t)}{dt} = -x(t) + v(t) + \sigma\eta(t)$$

similar but assume a decay or leakage in the accumulation process (or friction in brownian motion). Has a tendency of the walk to move back towards a central location, with a greater attraction when the process is further away from the center.

Both the Wiener and Ornstein-Ulhenbeck process have applications throughout mathematics and physics.

Drift Diffusion Model of Decision Making

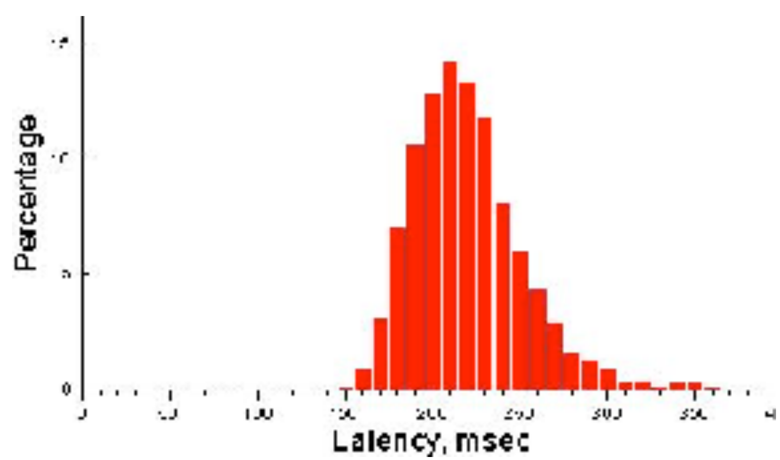
- all parameters will affect distribution of reaction times (RTs) (correct and/or errors)



Drift Diffusion model of Decision Making

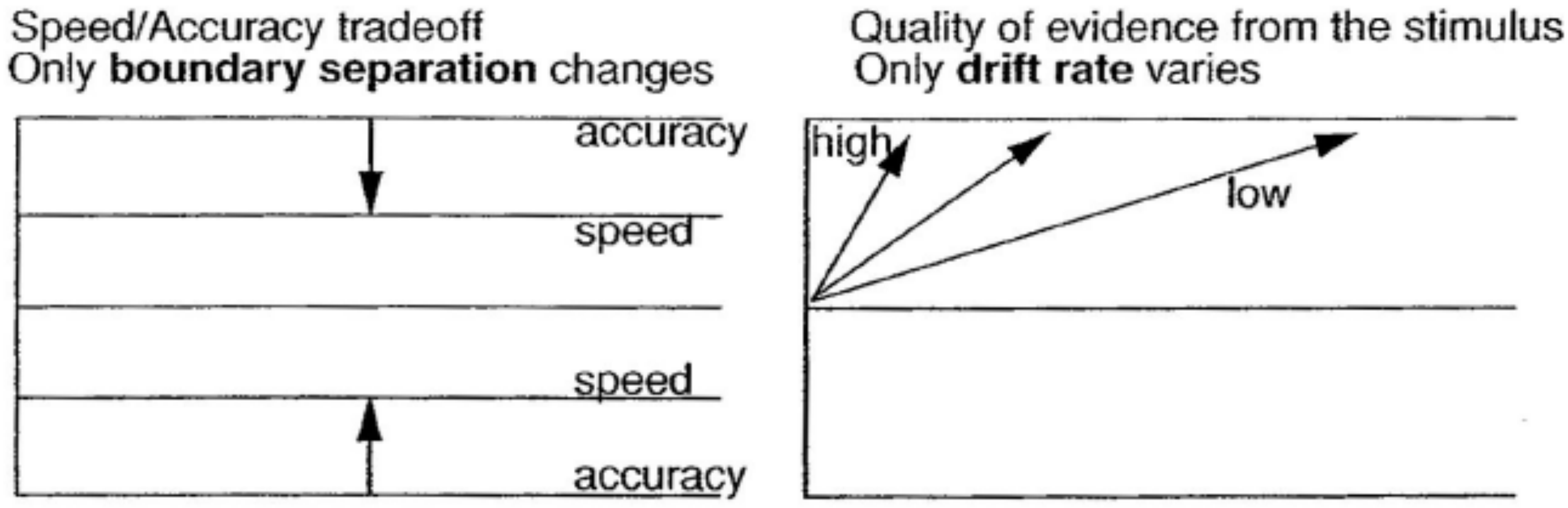
- many variants (discrete time, continuous time, leaky integration)
- These models have been compared systematically and shown to successfully account for [Smith & Ratcliff, 2004]:

- Distribution of Reaction Times
- Speed-accuracy tradeoff: decreasing the boundary has the effect of increasing speed and decreasing accuracy.
- Error response RTs (sometimes error responses can be very quick..).



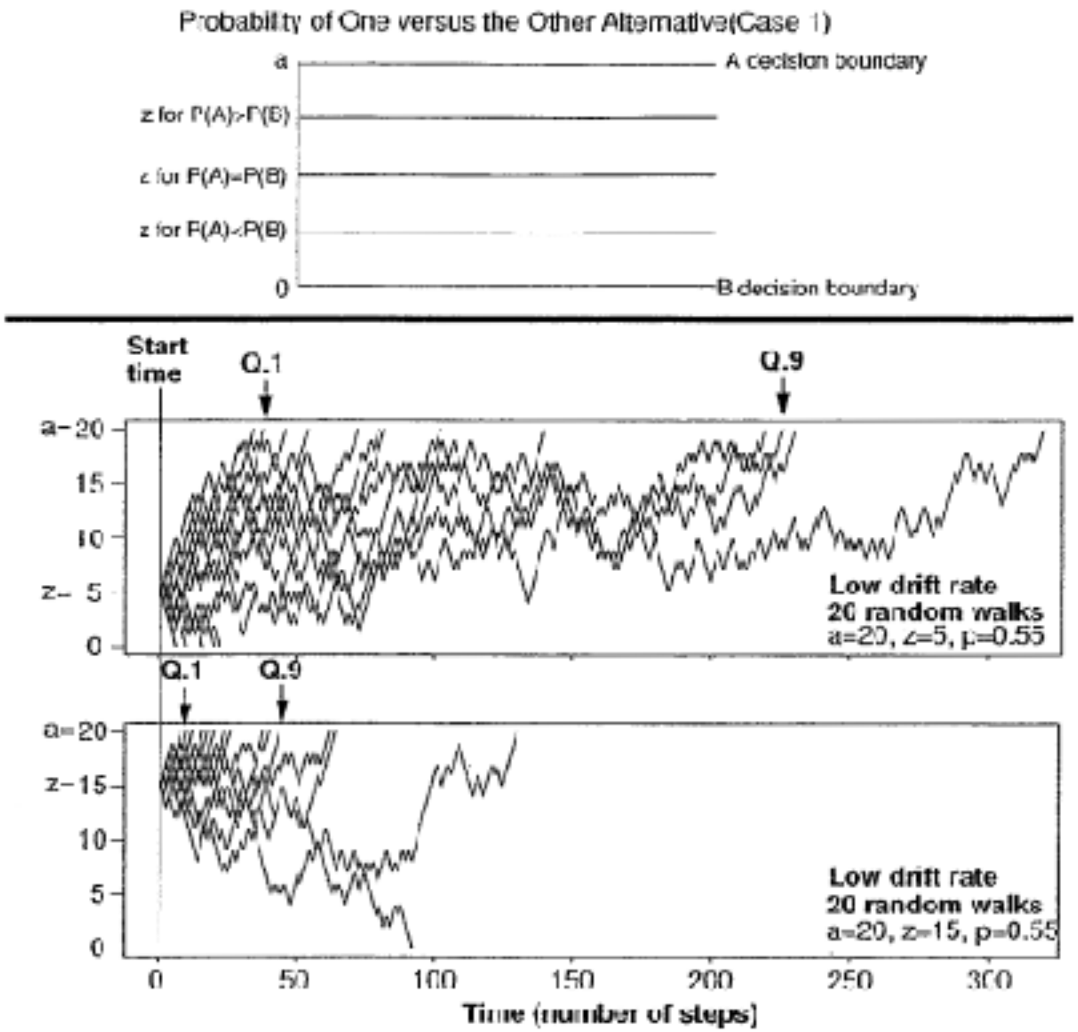
Drift Diffusion Model of Decision Making

- Drift rates changes with **difficulty** of conditions
- **speed-tradeoff accuracy** can be modelled by changing the boundary separation



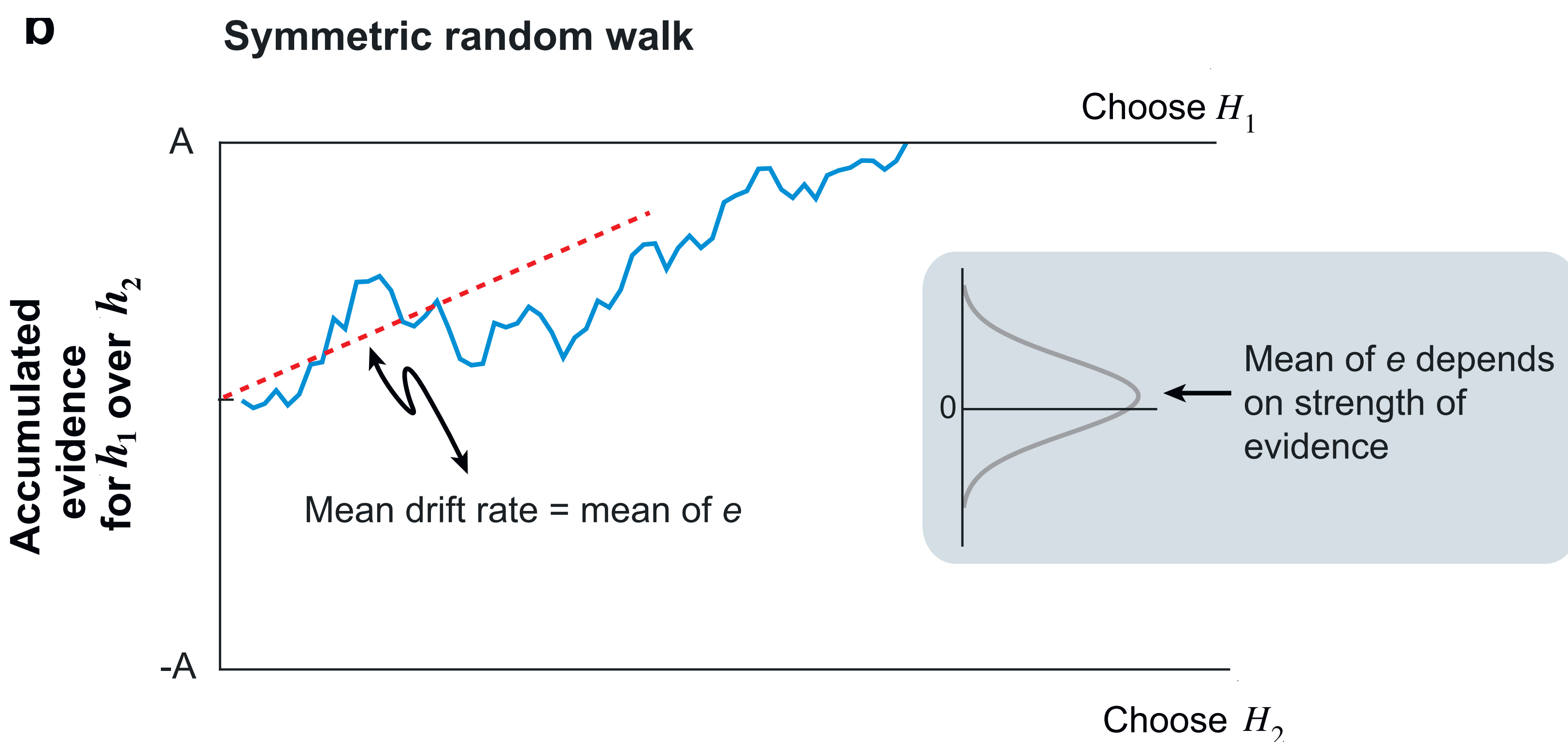
Drift Diffusion Model of Decision Making

- **Biases** can be introduced by changing the starting point.



Drift Diffusion Model of Decision Making

- Equivalence can be shown between DDM and sequential probability ratio test.
- The Decision Variable is the cumulated sum of the evidence. The bounds represent the stopping rule.
- If mean drift rate is log likelihood ratio, then this model = sequential probability ratio test.



Psychological Review
2006, Vol. 113, No. 4, 700–765

0033-295X/06/$12.00 DOI: 10.1037/0033-295X.113.4.700

The Physics of Optimal Decision Making: A Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks

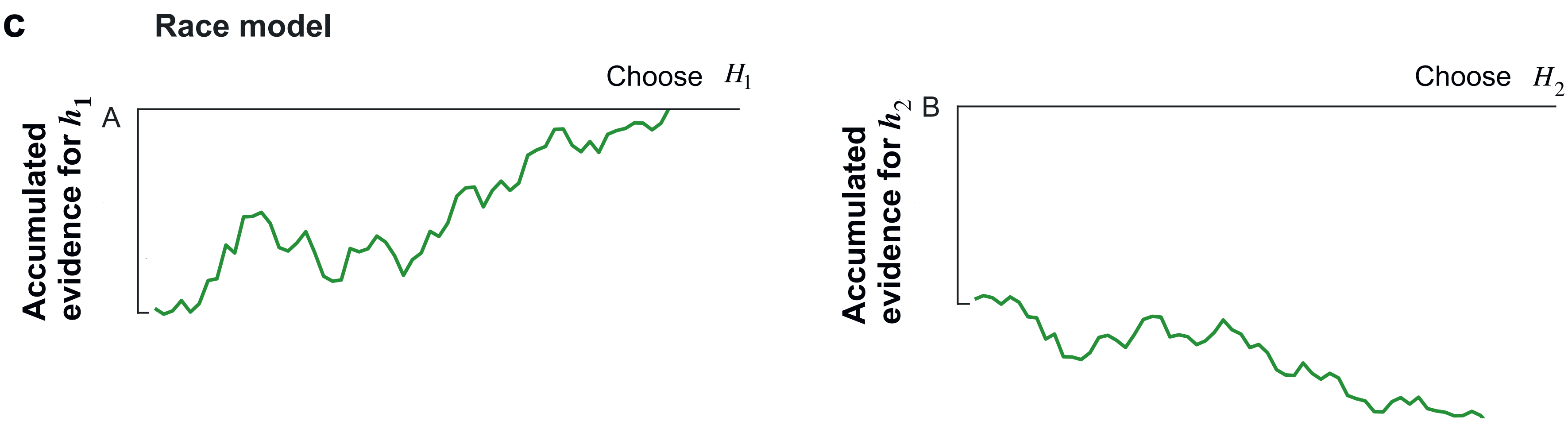
Rafal Bogacz, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D. Cohen
Princeton University

In this article, the authors consider optimal decision making in two-alternative forced-choice (TAFC) tasks. They begin by analyzing 6 models of TAFC decision making and show that all but one can be reduced to the drift diffusion model, implementing the statistically optimal algorithm (most accurate for a given speed or fastest for a given accuracy). They prove further that there is always an optimal trade-off between speed and accuracy that maximizes various reward functions, including reward rate (percentage of correct responses per unit time), as well as several other objective functions, including ones weighted for accuracy. They use these findings to address empirical data and make novel predictions about performance under optimality.

Keywords: drift diffusion model, reward rate, optimal performance, speed–accuracy trade-off, perceptual choice

Race Model

- Another variant is the race model
- Two or more decision processes represent the accumulated evidence for each alternative.



- Different properties

DDM as a tool in Computational Psychiatry

- The DDM can be used as a tool to compare groups.
- Because it is a unified model of speed and accuracy, it can be more sensitive than looking at RTs or % correct alone.

- Example: Do participants with anxiety respond differently to threatening words?

Emotion
2010, Vol. 10, No. 5, 662–677

1528-3542/10/$12.00 DOI: 10.1037/a0019

Anxiety Enhances Threat Processing Without Competition Among Multiple Inputs: A Diffusion Model Analysis

Corey N. White, Roger Ratcliff, Michael W. Vasey, and Gail McKoon
Ohio State University

Enhanced processing of threatening information is a well established phenomenon among high-anxious individuals. This effect is most reliably shown in situations where 2 or more items compete for processing resources, suggesting that input competition is a critical component of the effect. However, it could be that there are small effects in situations without input competition, but the dependent measures typically used are not sensitive enough to detect them. The present study analyzed data from a noncompetition task, single-string lexical decision, with the diffusion model, a decision process model that provides a more direct measure of performance differences than either response times or accuracy alone. The diffusion model analysis showed a consistent processing advantage for threatening words in high-anxious individuals, whereas traditional comparisons showed no significant differences. These results challenge the view that input competition is necessary for enhanced threat processing. Implications for theories of anxiety are discussed.

Avocad

Word or (a)
non word
(l)?

Subvi

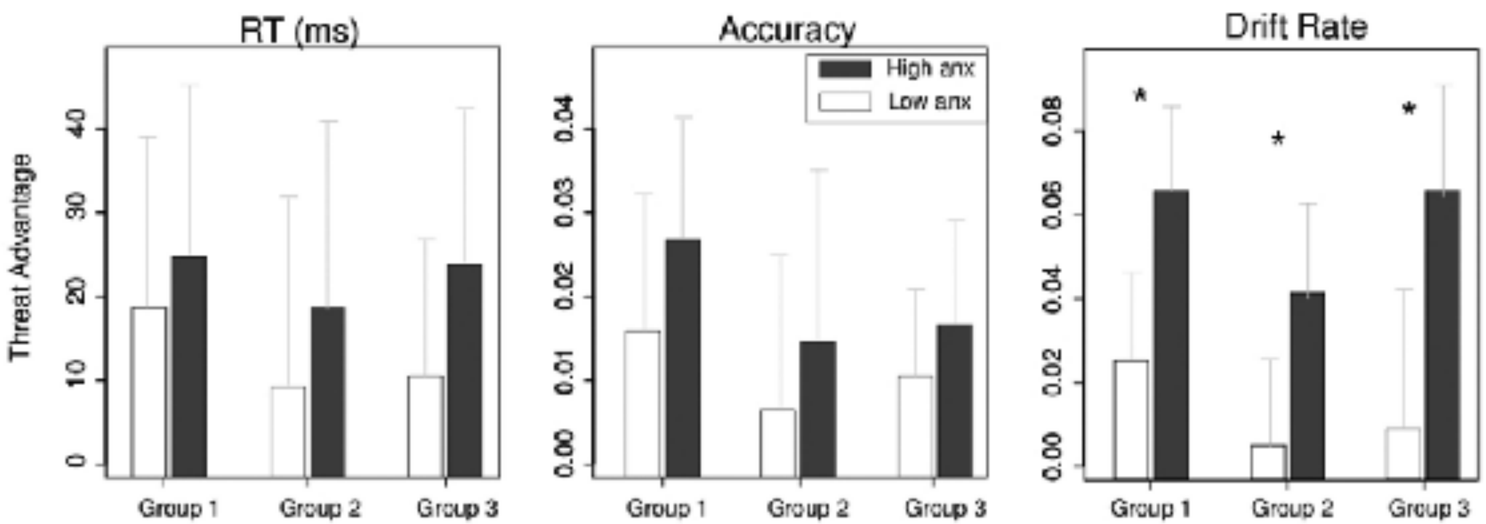
Word or (a)
non word
(l)?

death

Word or (a)
non word
(l)?

Drift Diffusion Model of Decision Making

- The DDM is fit to individual participants
- Quality of fit is assessed by how it fits the RT distributions
- Participants with high anxiety had larger drift rates for threatening compared to nonthreatening words whereas participants with low anxiety did not.
- Suggests enhanced processing of threatening words for participants with high anxiety



(Intermediate) Conclusions

- A decision = process that weights priors, evidence, and value to generate a commitment
- Signal detection theory and sequential analysis provide a theoretical framework for understanding how optimal decisions can be made.
- Idea: a decision variable (~logLR) is compared to a threshold
- DDM, invented as a phenomenological framework to describe human behaviour, % accuracy and RTs, found to be equivalent.
- DDM used as a tool in computational psychiatry.
- In practice nowadays, hierarchical DDM (hDDM) - toolbox using Bayesian hierarchical parameter estimation.

Further Readings

- CP section 2.2 (DDM)
- Page maintained by Ratcliff:

https://u.osu.edu/ratcliffmckoon/the-diffusion-model-for-non-specialists/

- hDDM: Wiecki, Sofer and Frank (2013):

http://ski.clps.brown.edu/hddm_docs/

- DDM tutorial: https://www.med.upenn.edu/longding1/javascript/DDM_LongDing.html