Making decisions quickly (part 2)

Computational Cognitive Science 2014

Dan Navarro

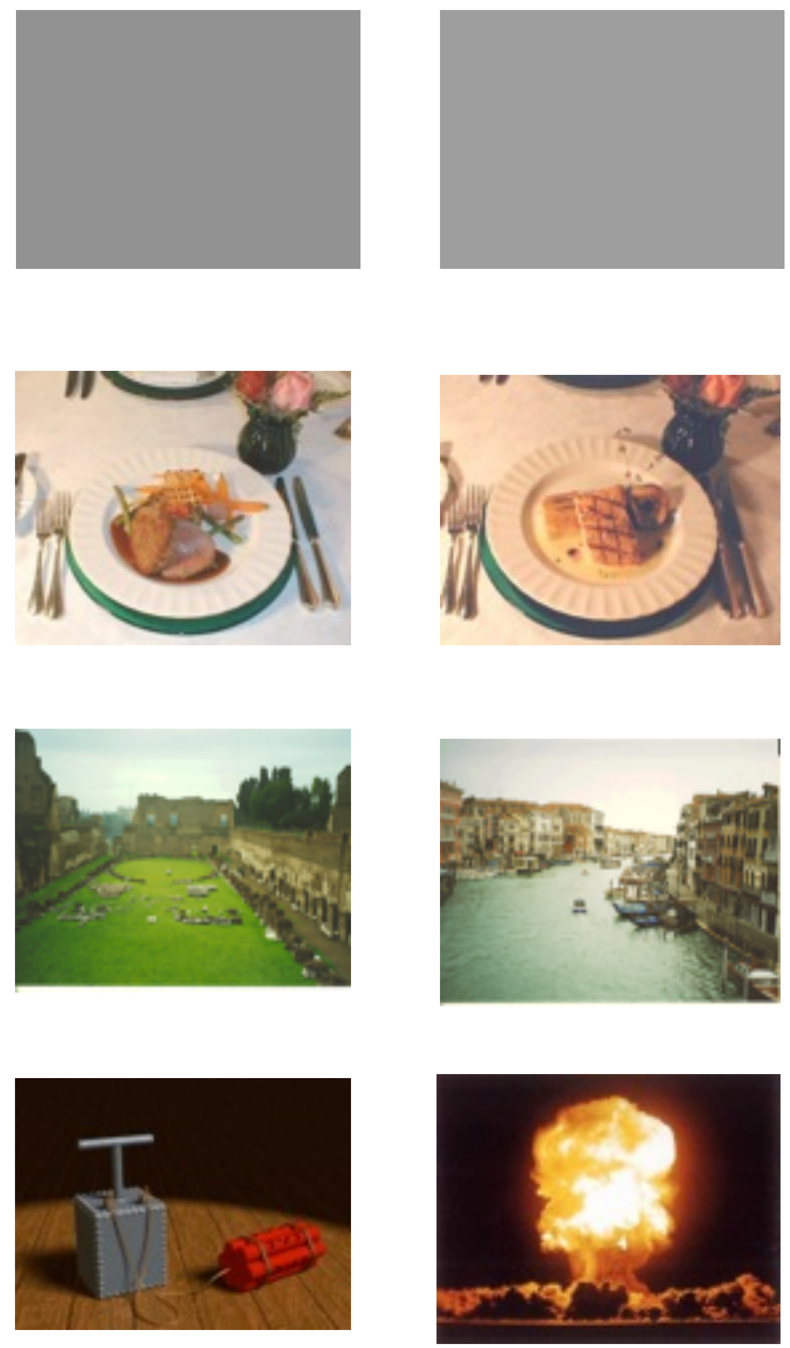
Overview of the lectures

- Last lecture:
  - Historical background: psychophysics
  - Introduction to signal detection theory
  - The utility of time and computation
  - Introduction to sequential sampling models

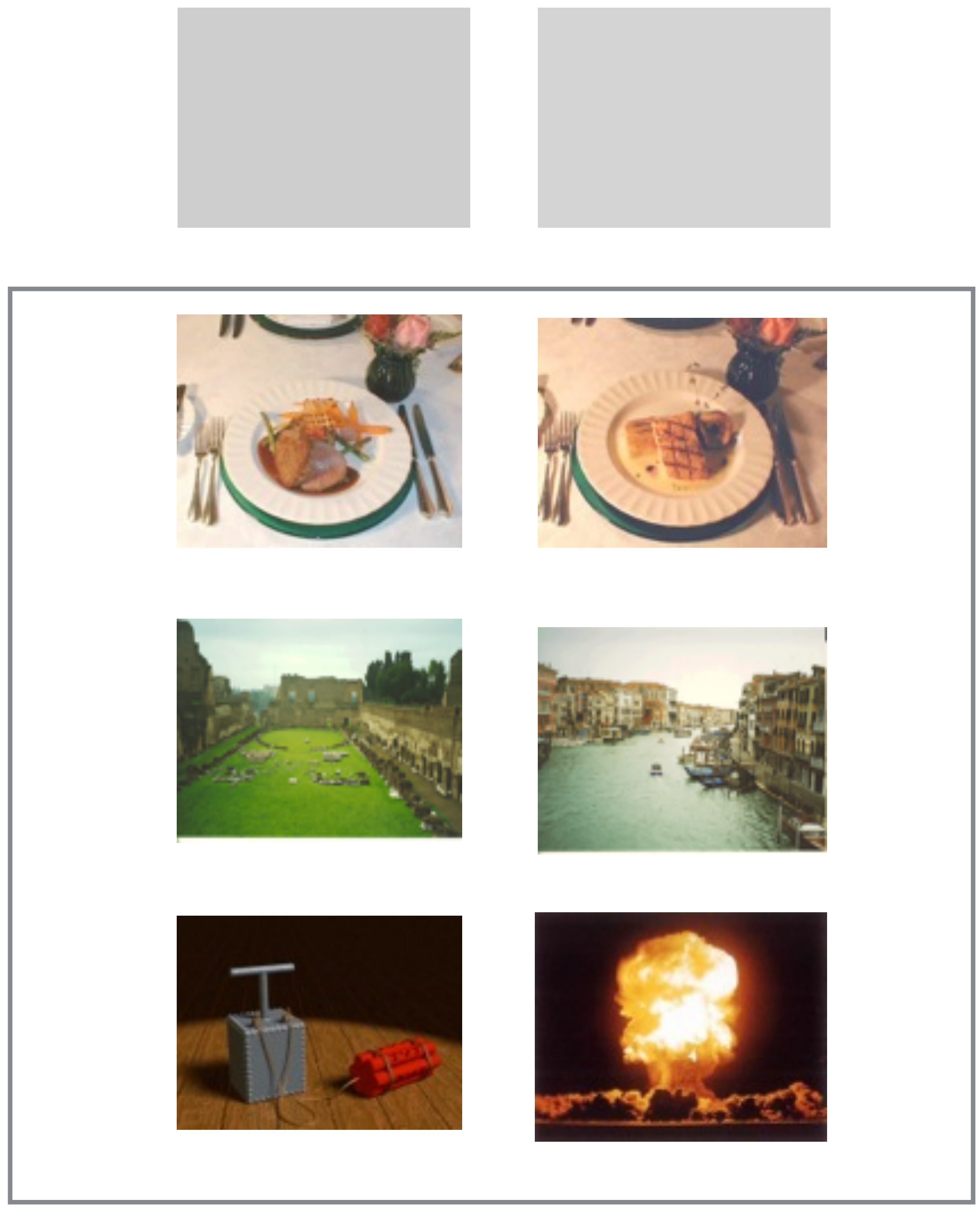
- This lecture
  - More on sequential sampling models
  - Applications of SSMs to cognitive science
  - Using SSMs in machine learning
  - Using SSMs in neuroscience

Quick review from last time

Many kinds of decisions

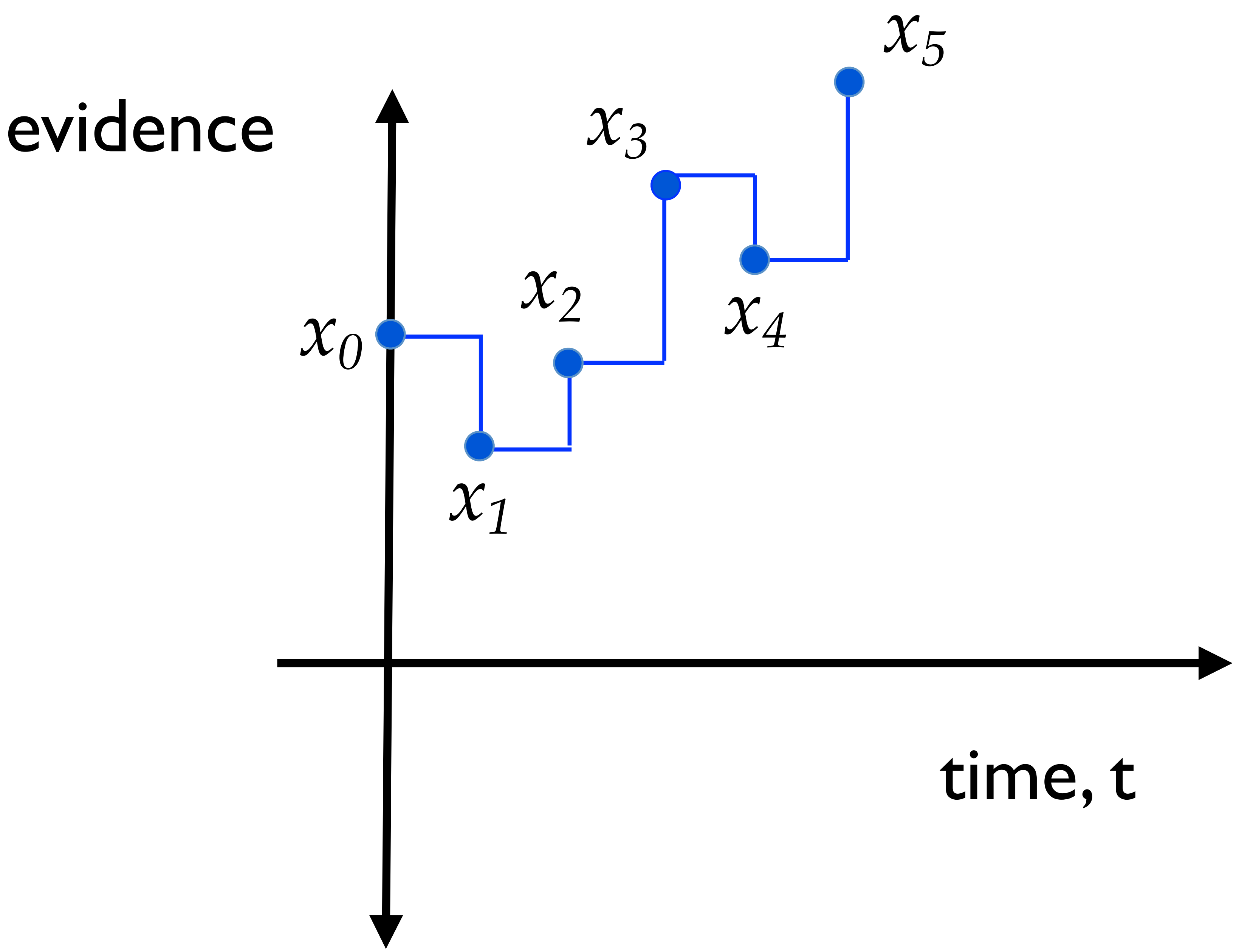


Complex decisions tell us a lot about utility/ cost functions (i.e. they don't exist)

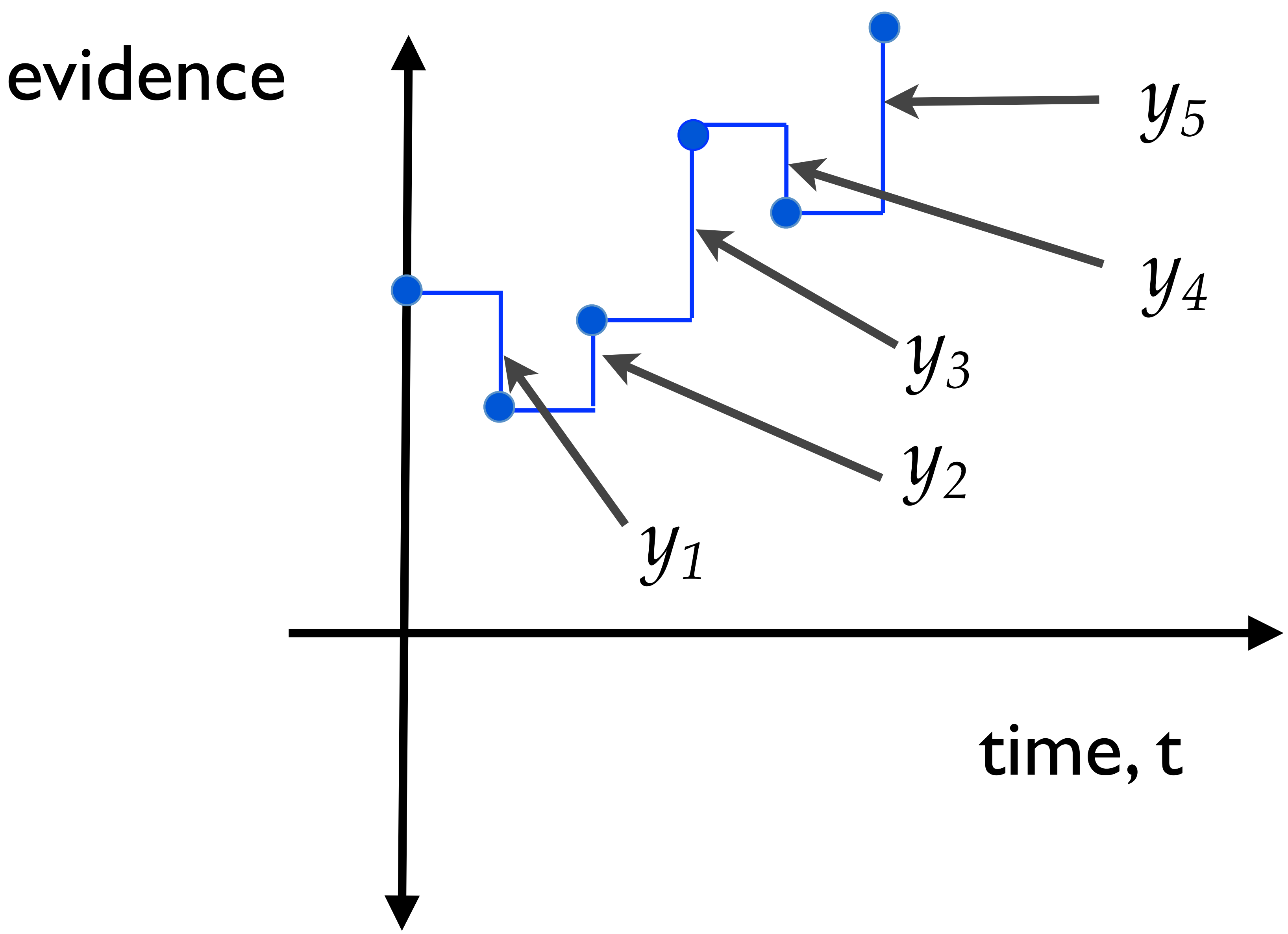


Simple decisions tell us a lot about basic information processing over <u>time</u>

A random walk over "evidence space"



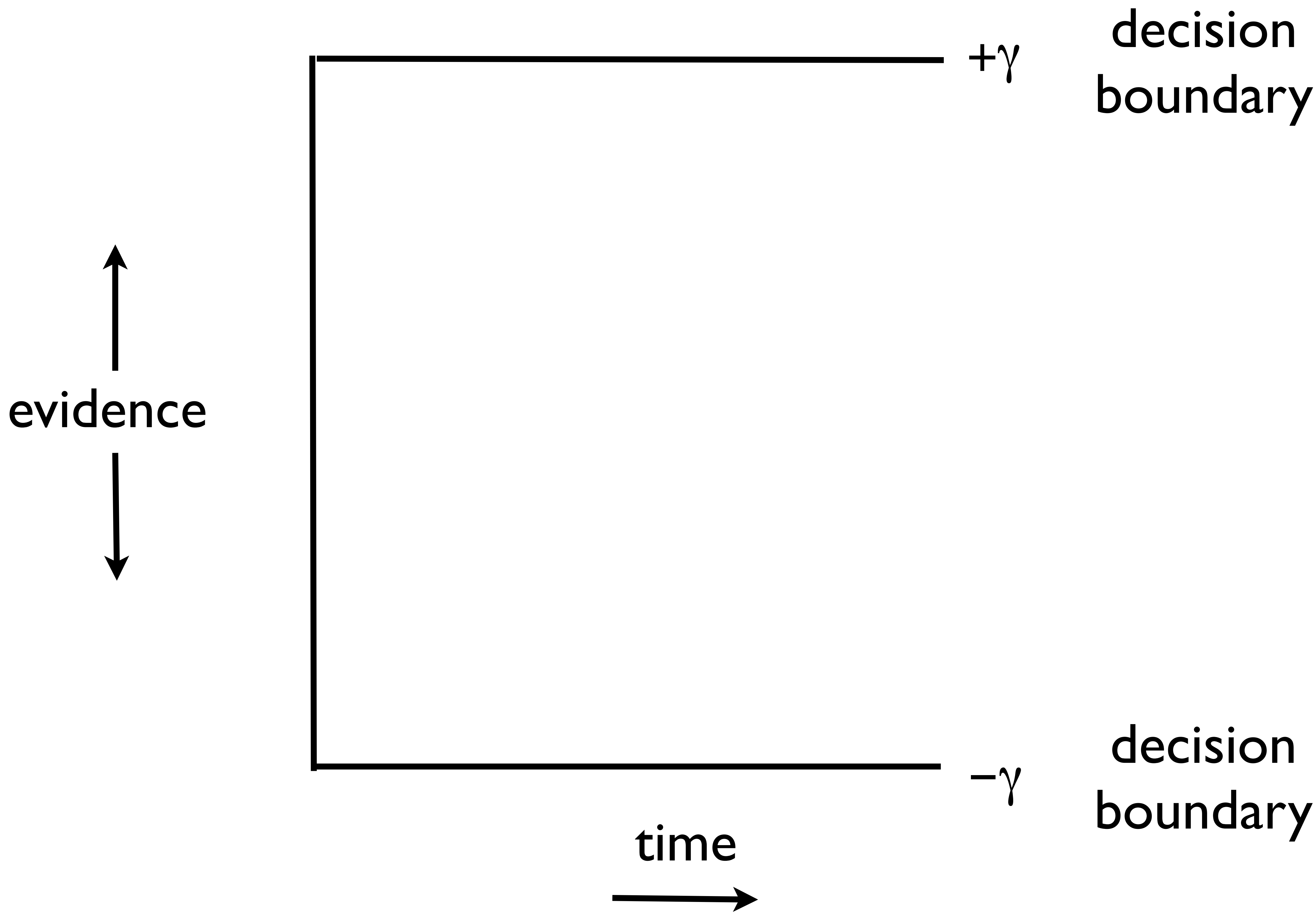
The size of each "step" corresponds to the evidence provided by a sample



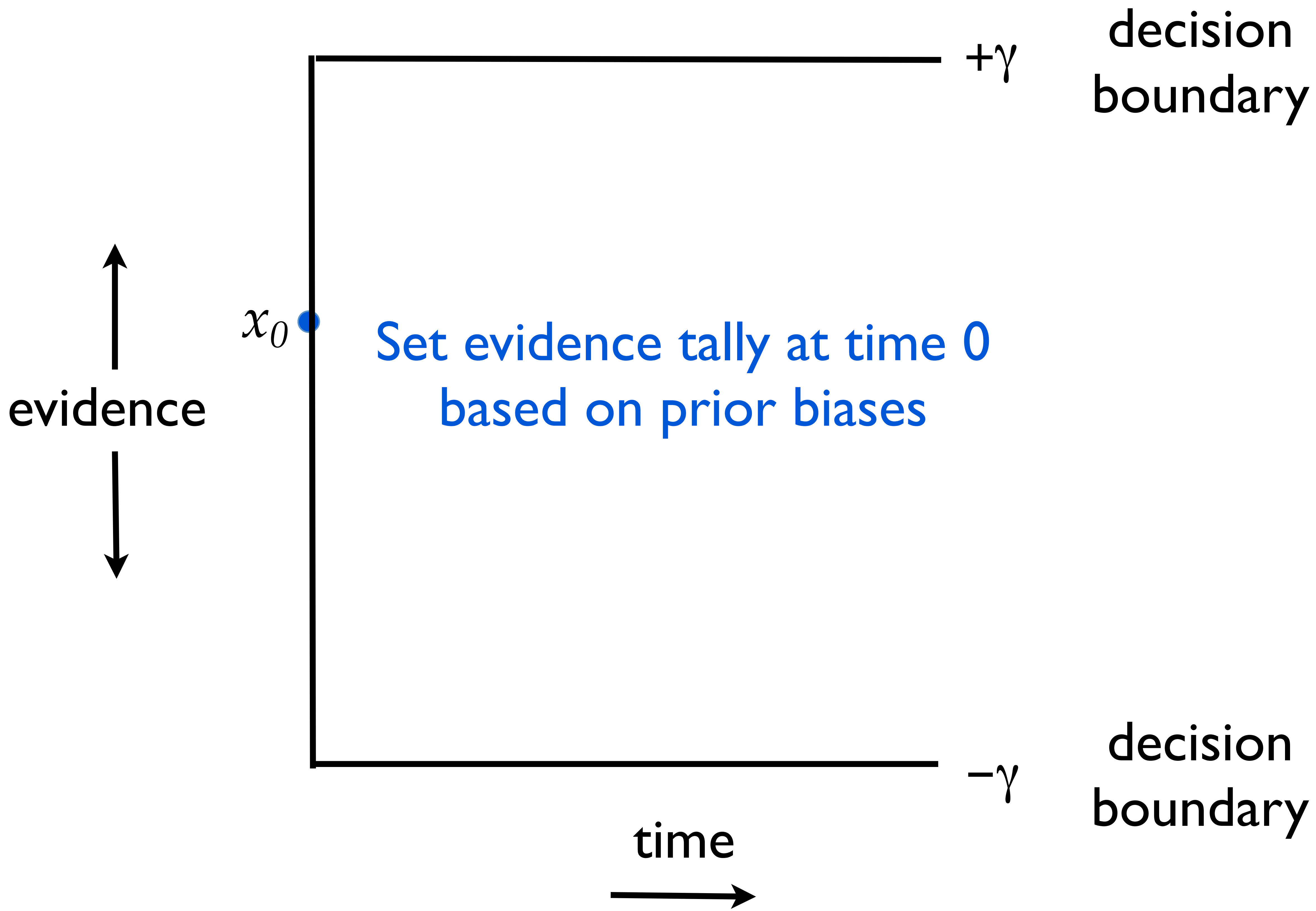
The random walk model for simple decisions

1. Time $t = 0$
2. Set x_0, based on your prior biases
3. Do while $|x_t| < \gamma$
   i. Time increments, $t = t+1$
   ii. Collect sensory sample s_t
   iii. Evaluate the log-odds for that sample, y_t
   iv. Increment evidence tally, $x_t = x_{t-1} + y_t$
4. If $x_t \geq \gamma$, choose option A
5. If $x_t \leq -\gamma$, choose option B

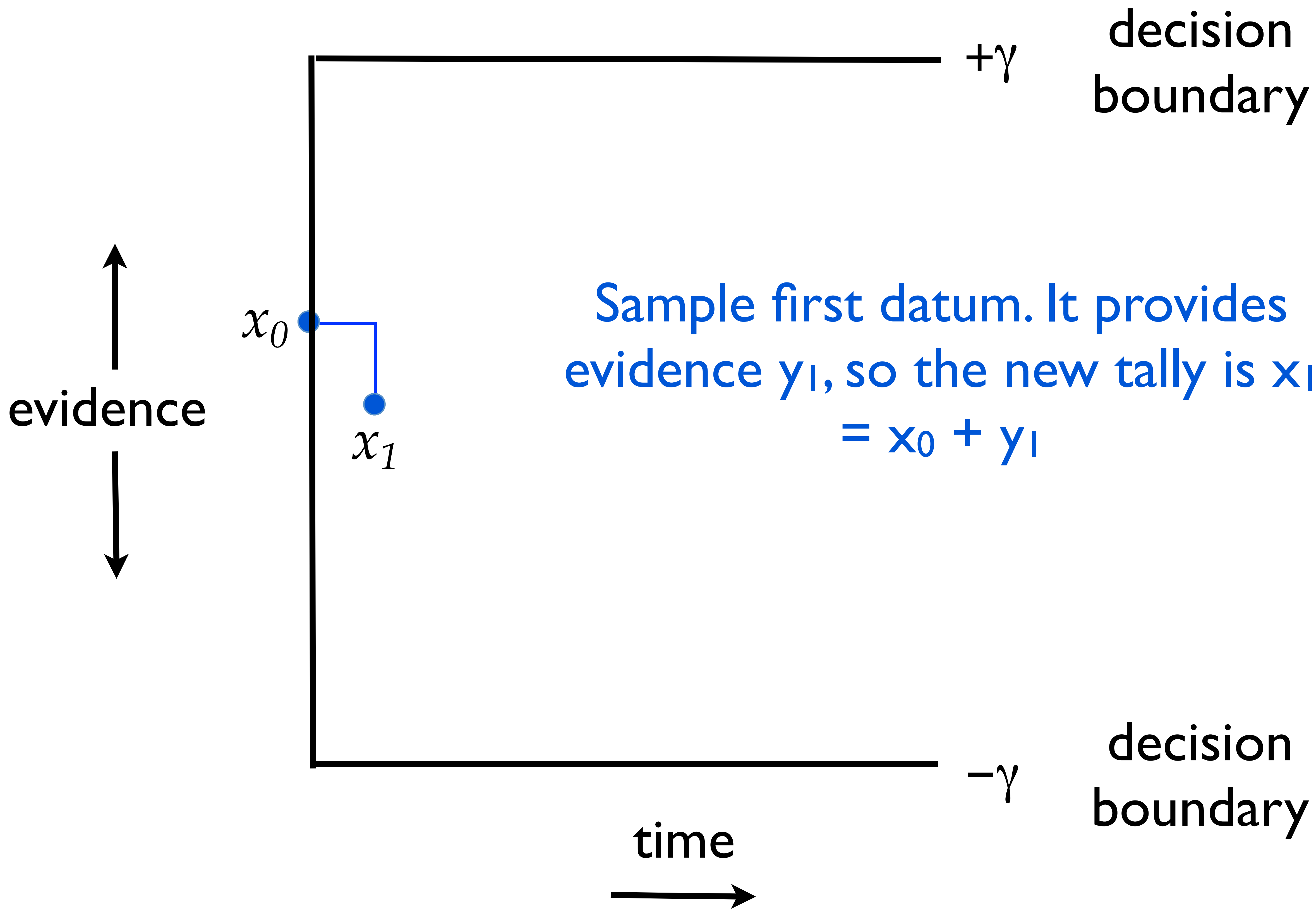
The random walk model



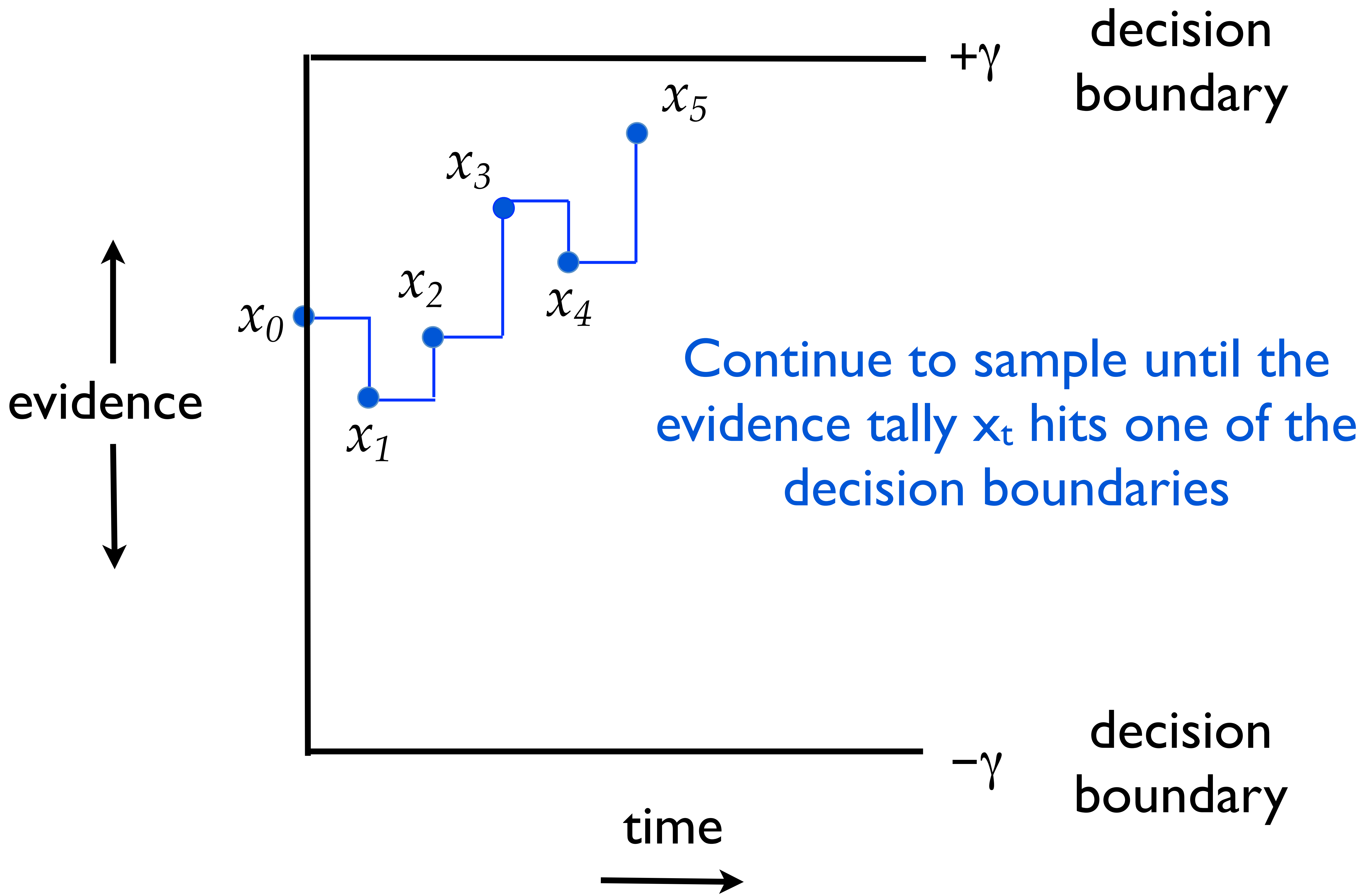
The random walk model



The random walk model



The random walk model



RANDOM WALK MODEL

set γ based on your tolerance for errors
set x_0 based on your prior beliefs
set time $t=0$
while $|x_t| < \gamma$
    $t=t+1$
    draw y_t from the **information function**
    increment beliefs $x_t = x_{t-1} + y_t$
make decision:
    $r=1$ if $x_t \geq \gamma$ (i.e. "choose A")
    $r=0$ if $x_t \leq \gamma$ (i,e,"choose B")
output r and t

"BERNOULLI" INFORMATION FUNCTION

input p

generate $u \sim \mathrm{Uniform}([0,1])$

if $u \le p$

  $y = 1$

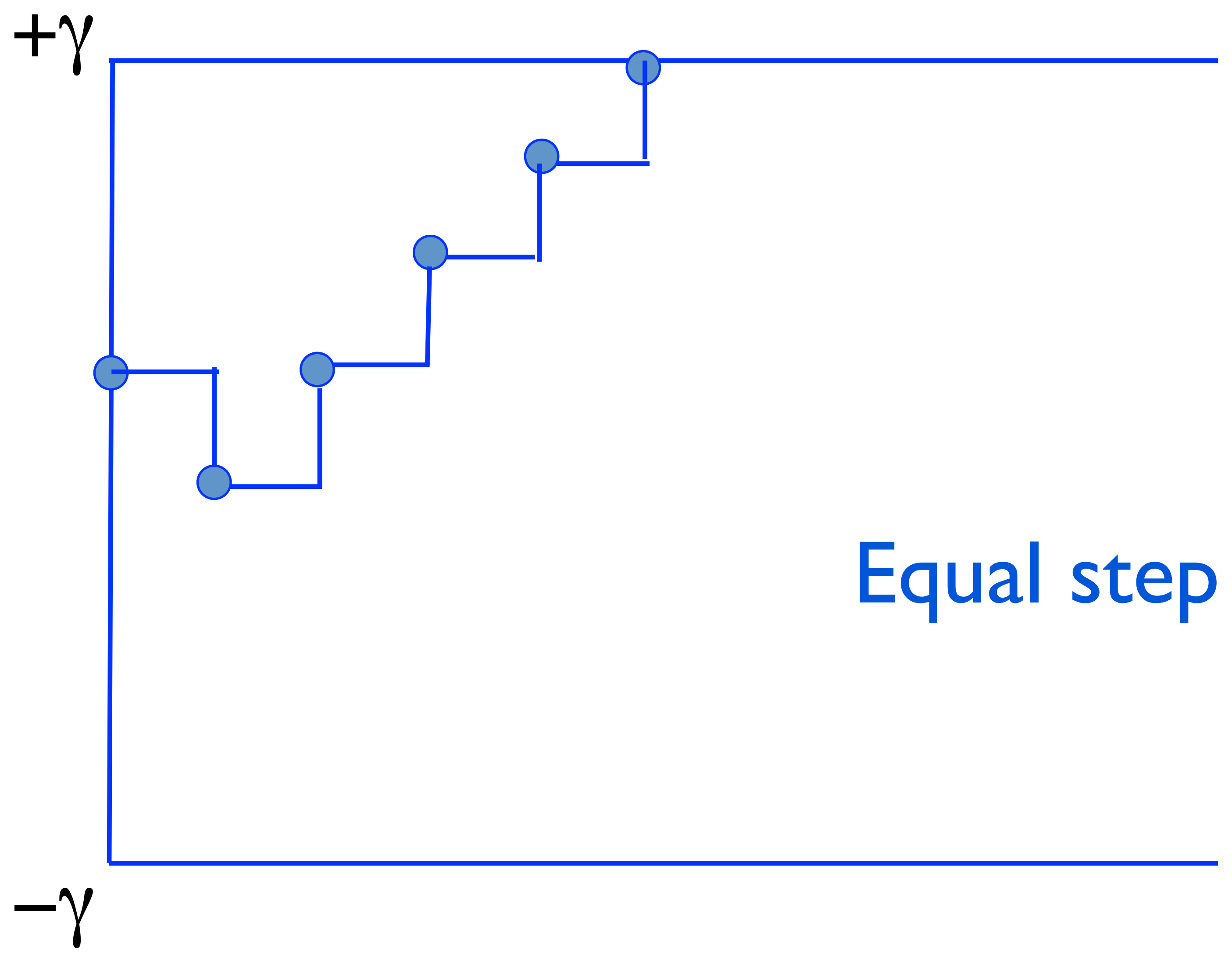
else

  $y = -1$

output y

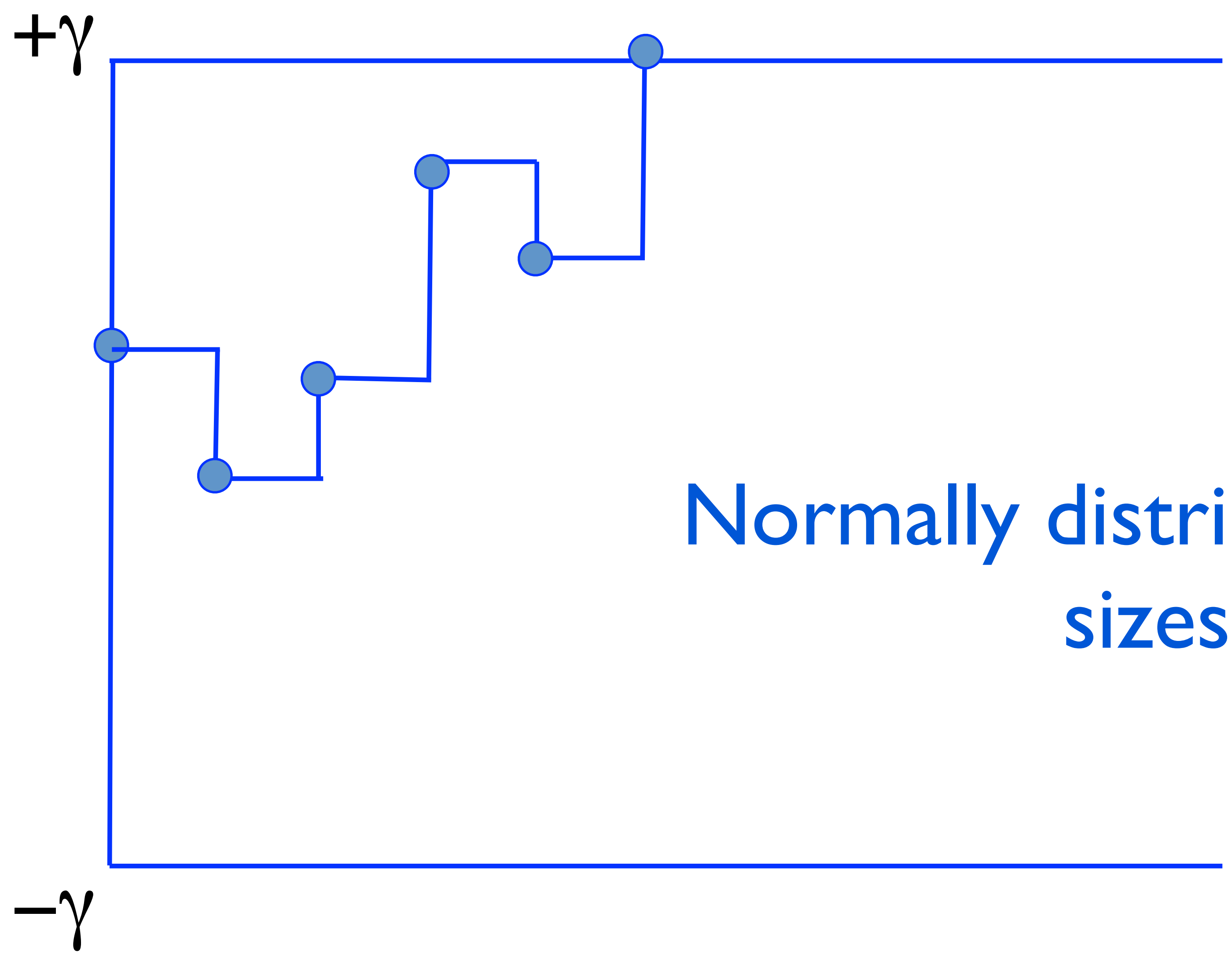
Step up with probability p,
step down with probability 1-p

A "Bernoulli" random walk model generates paths that look like this one:



Equal step sizes

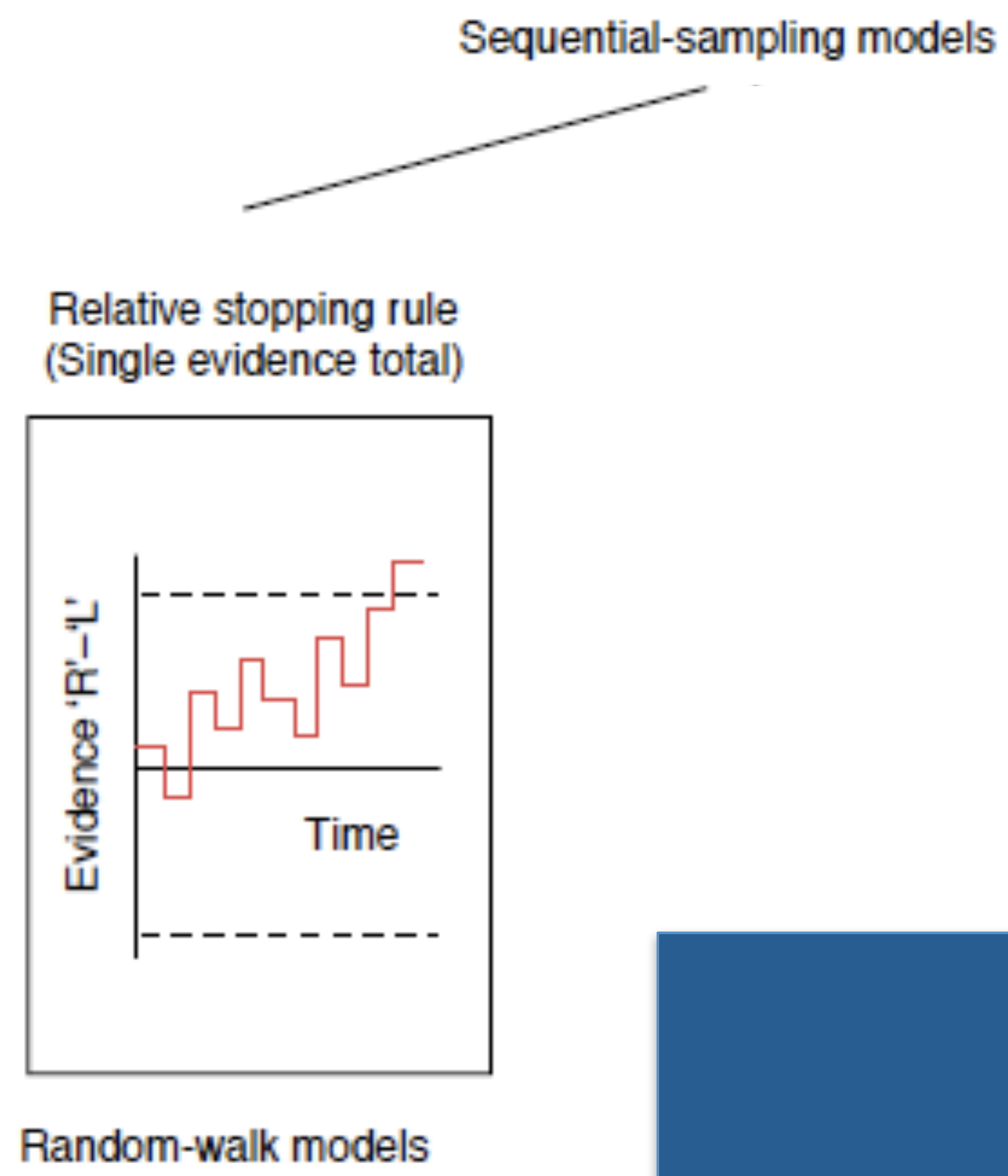
A "Gaussian" random walk model generates paths that look like this one:



Normally distributed step sizes

Demo: ssm.R

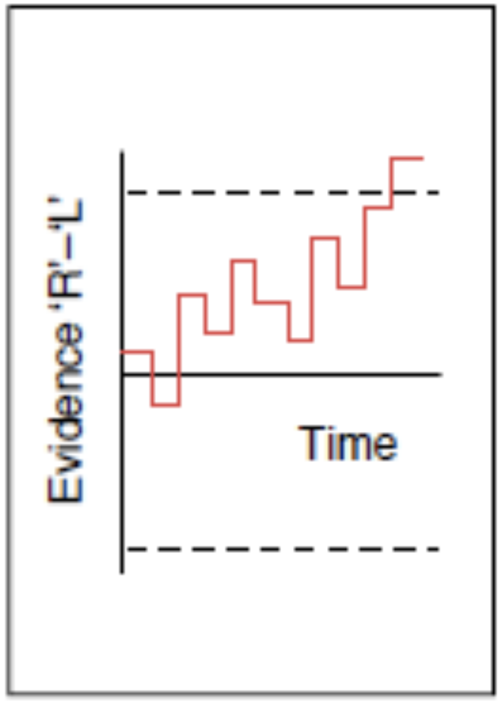
There are a lot of variations on SSMs…
An incomplete taxonomy



Random walks (the kind of SSM we've seen so far) are just one example of a sequential sampling model

Sequential-sampling models

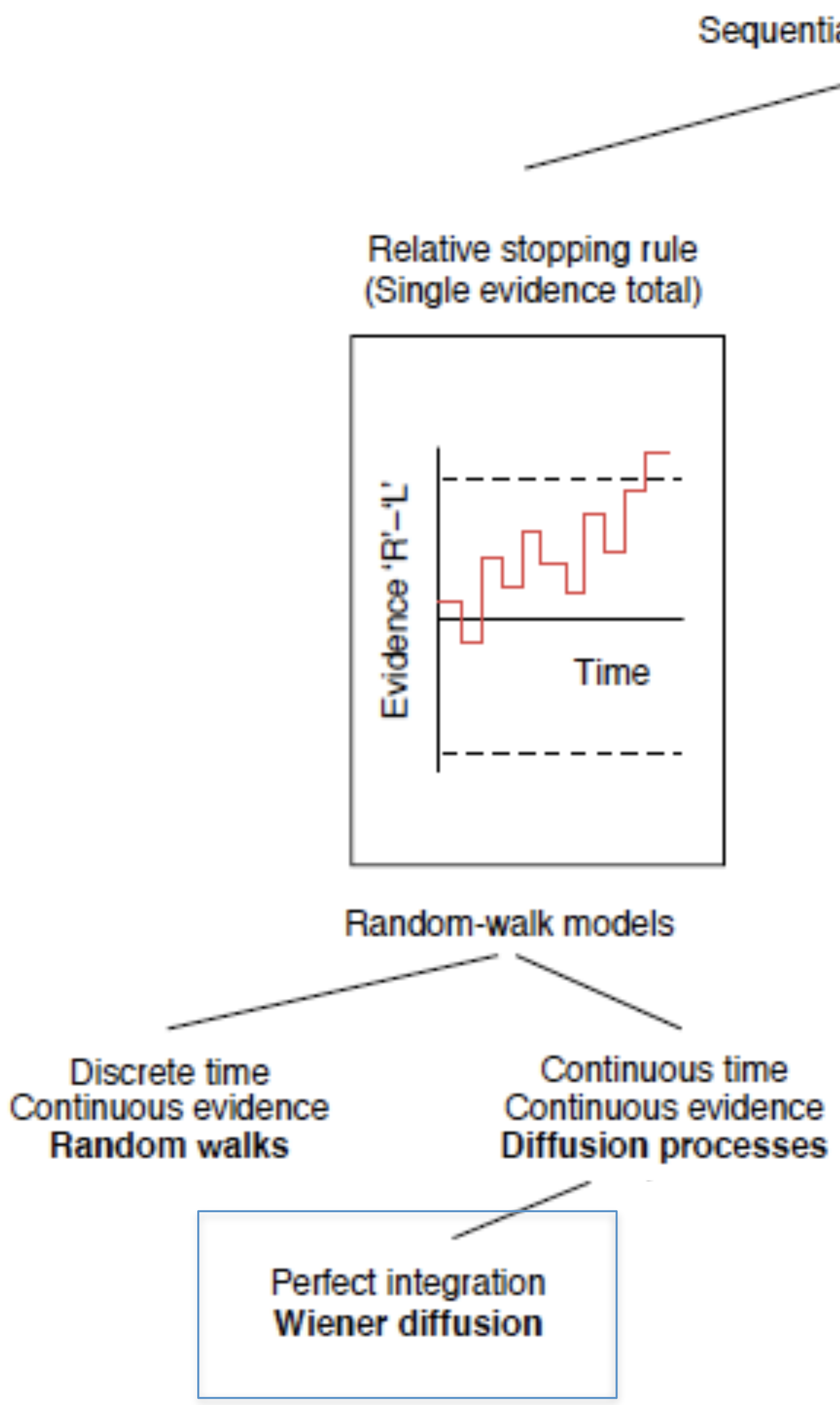
Relative stopping rule
(Single evidence total)



Random-walk models

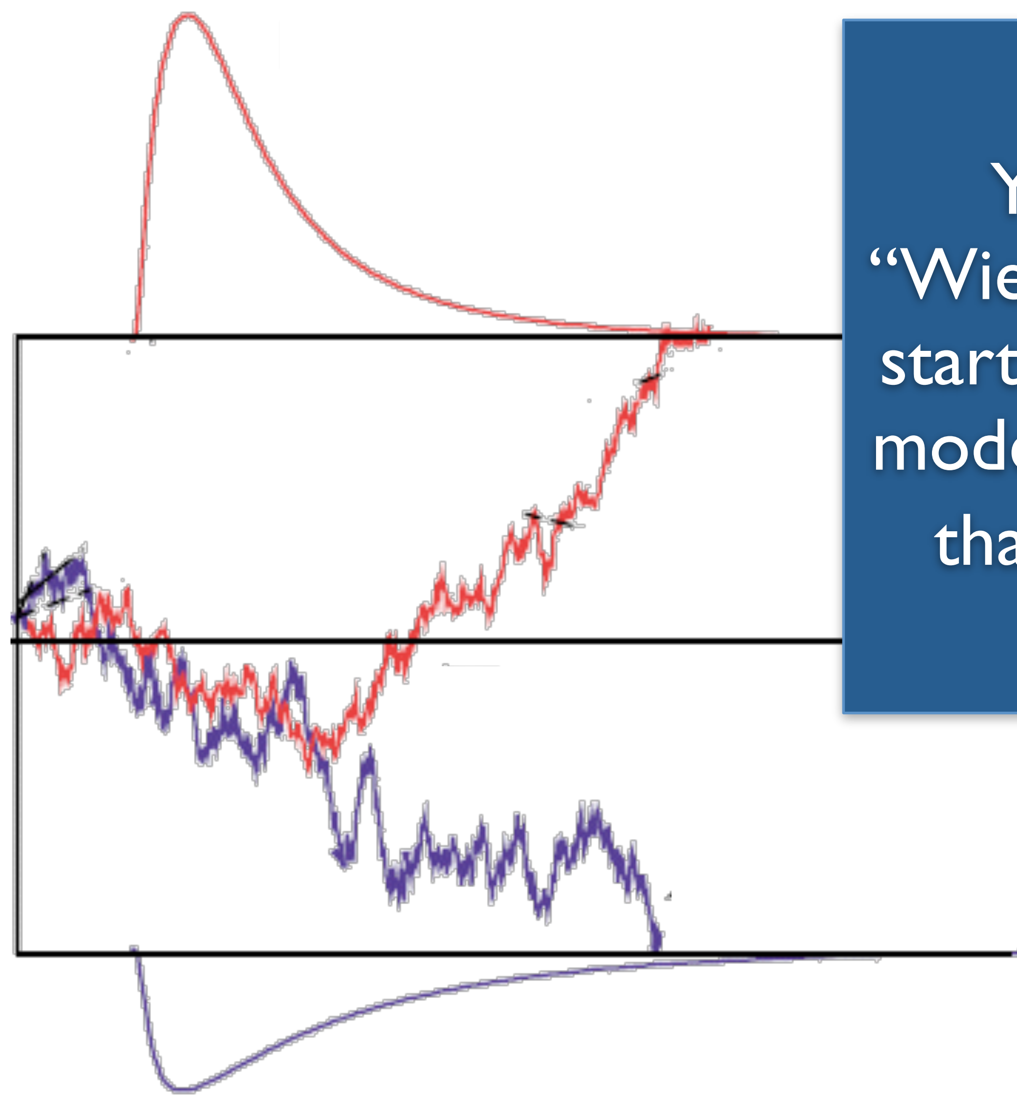
Discrete time
Continuous evidence
Random walks

The particular random walk model I just derived operates in discrete time, but with continuous-valued evidence



Others operate in continuous time, with continuous evidence

“The assumption of small steps”



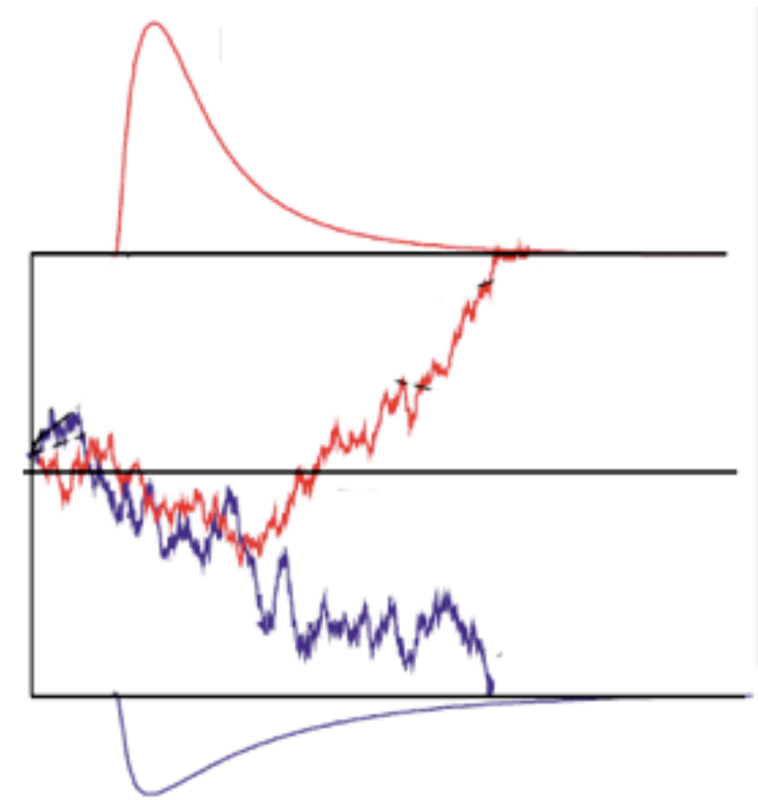
You can produce the “Wiener diffusion model” by starting with a random walk model, and taking limits such that $\Delta t \rightarrow 0$ and $\Delta x \rightarrow 0$

(Comments)

1. Feller (1968) derives expressions for these distributions. The answer involves infinite series, but computations can be made very fast.

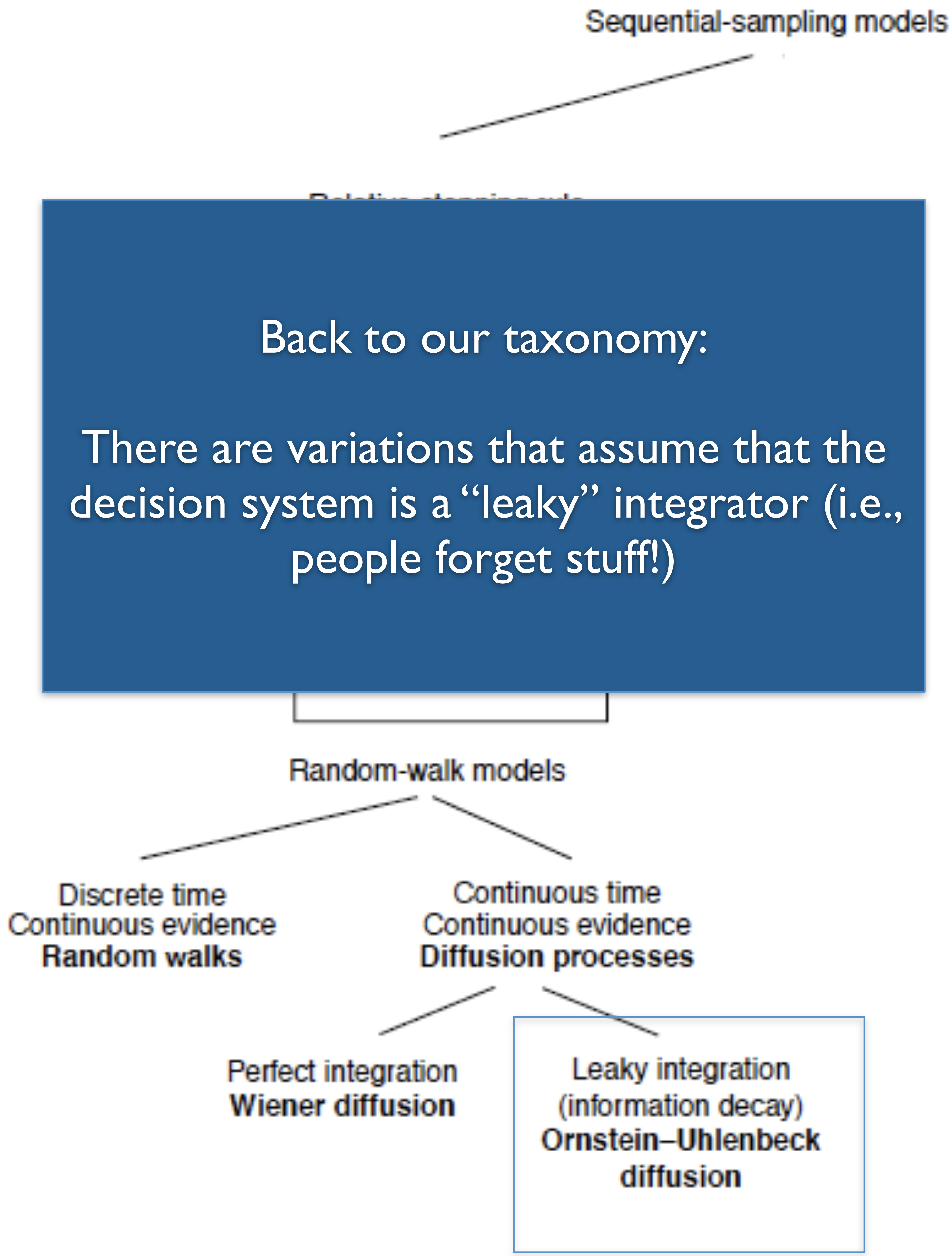
2. Navarro & Fuss (2009) and Blurton et al (2012) provide the analytic results that let you do this

3. The RWeiner package in R implements it

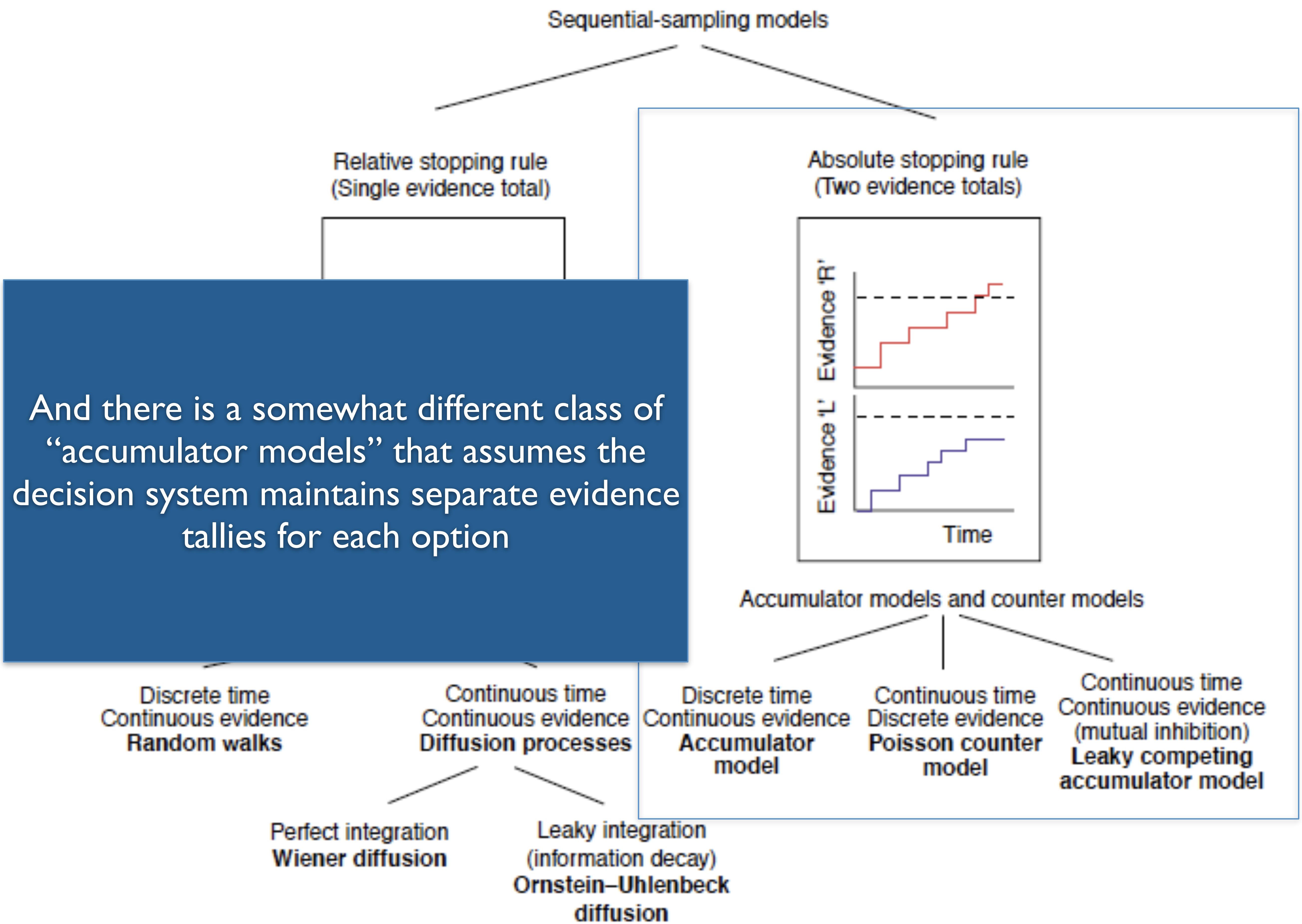


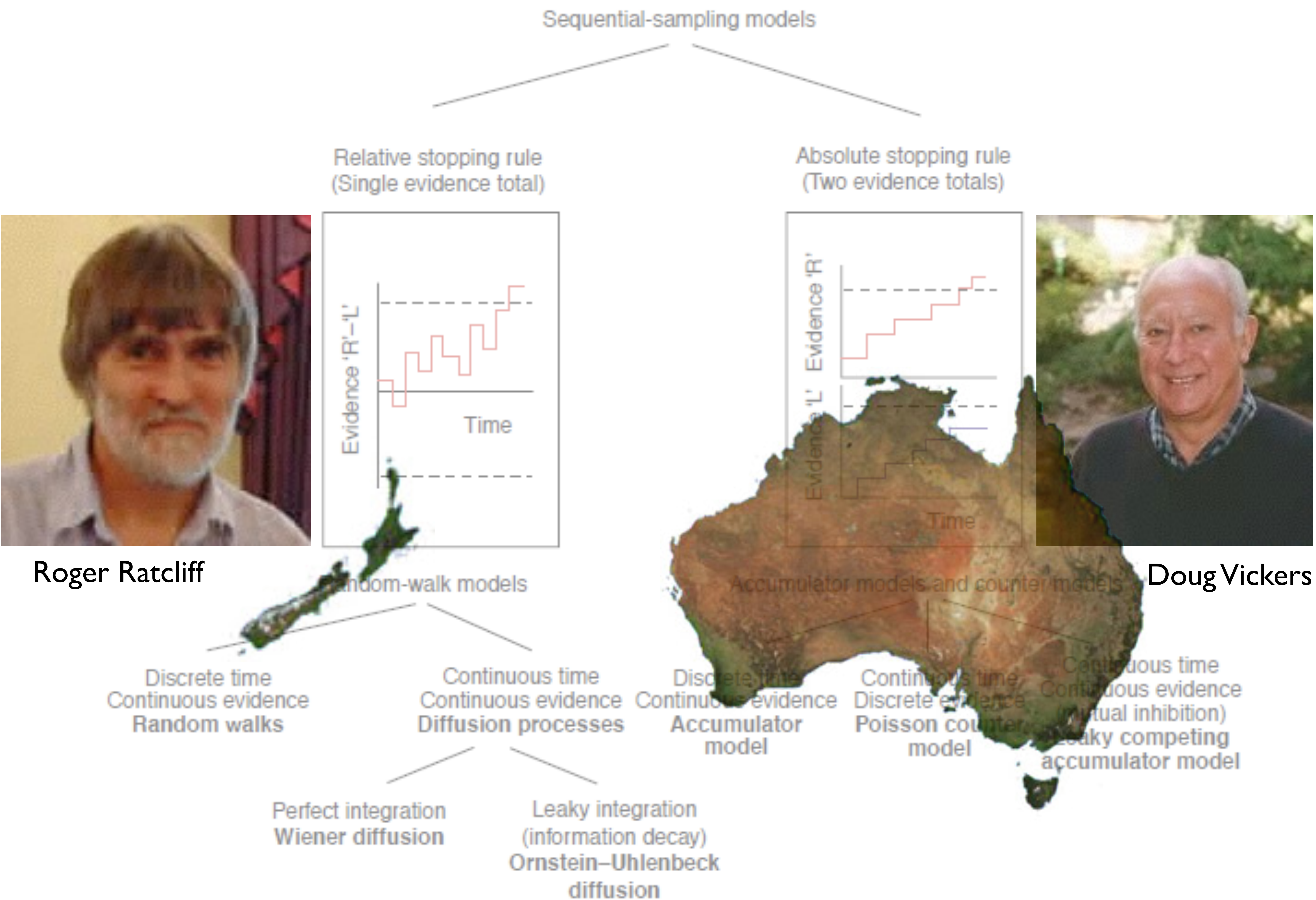
Back to our taxonomy:

There are variations that assume that the decision system is a “leaky” integrator (i.e., people forget stuff!)



And there is a somewhat different class of "accumulator models" that assumes the decision system maintains separate evidence tallies for each option





Sequential-sampling models
Relative stopping rule
(Single evidence total)
Absolute stopping rule
(Two evidence totals)
Evidence 'R'–'L'
Time
Time
Roger Ratcliff
Doug Vickers
Discrete time
Continuous evidence
Random walks
Continuous time
Continuous evidence
Diffusion processes
Accumulator
model
model
accumulator model
Perfect integration
Wiener diffusion
Leaky integration
(information decay)
Ornstein–Uhlenbeck
diffusion

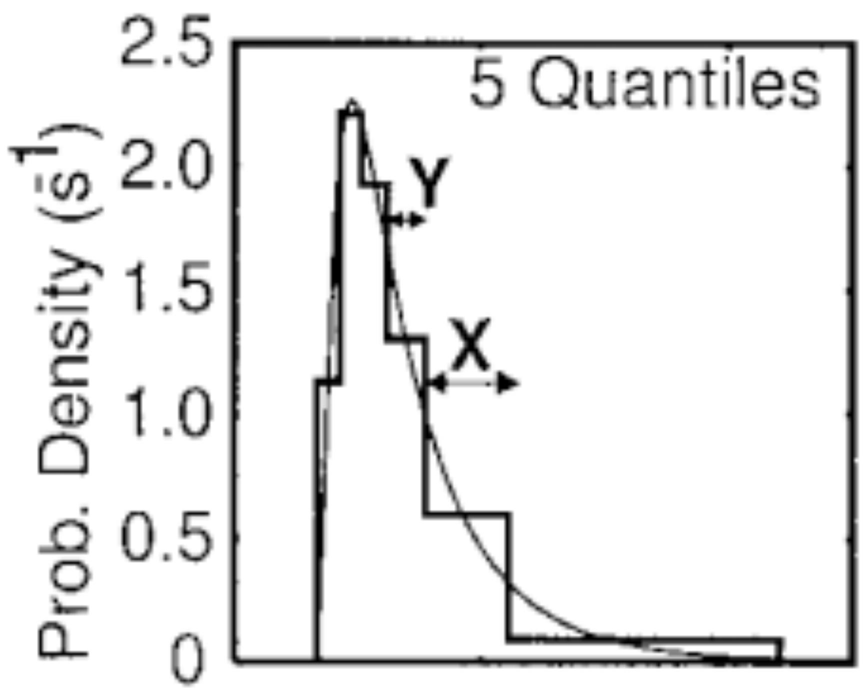
How well do these models work?

The SSM framework is very general

- When can an SSM be applied?
  - Originally the models were designed to handle simple perceptual choice problems
  - But they can be applied in ANY situation where humans make decisions that unfold over time
  - That is, they're intended to be a kind of "universal" front end that explains choice and RT in any task.

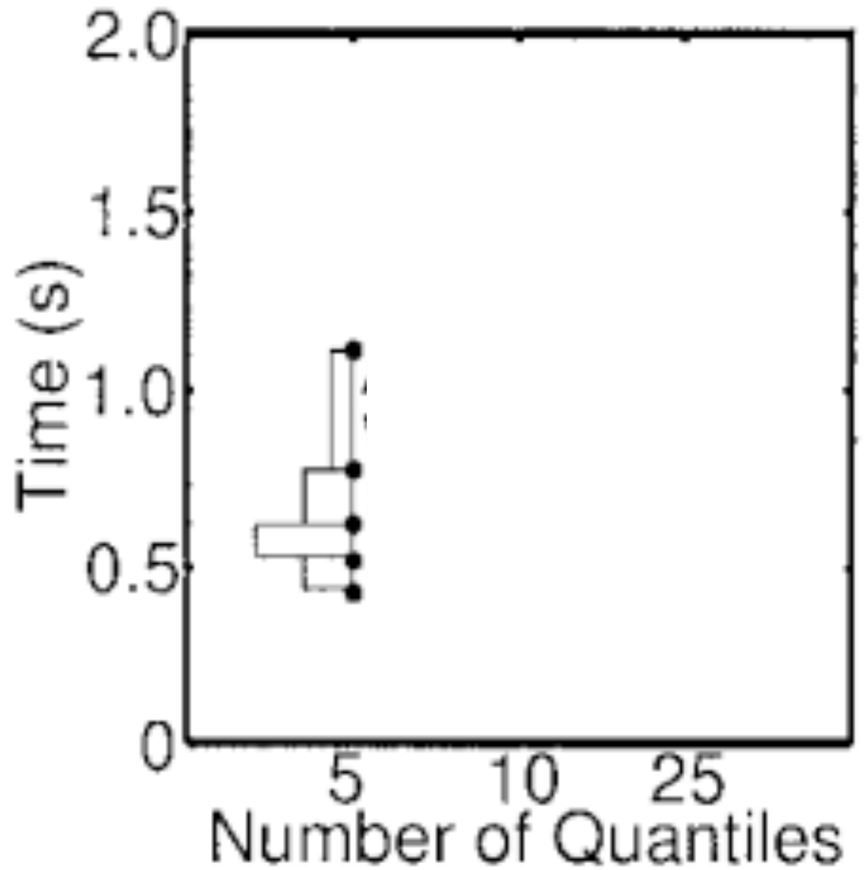
- Do they work?
  - Yes. Very, very well.
  - But first… quantile probability functions…

Quantile probability functions



For both models and humans, "cut" the data up into N separate "bins"... lowest bin contains all of the fastest RTs, highest bin contains the slowest RTs, etc (note: ignore the lowest 5% and highest 5%)

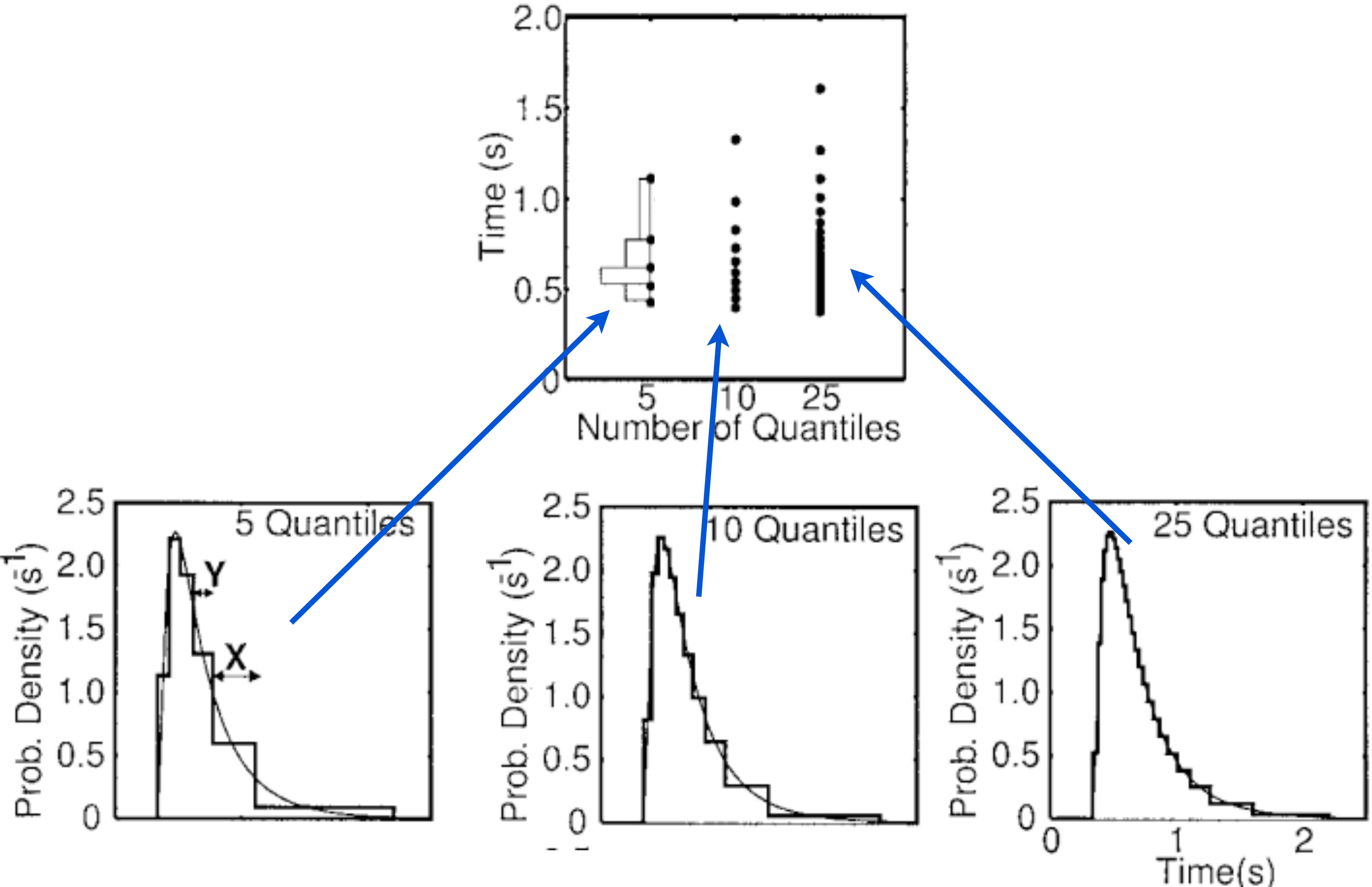
Quantile probability functions



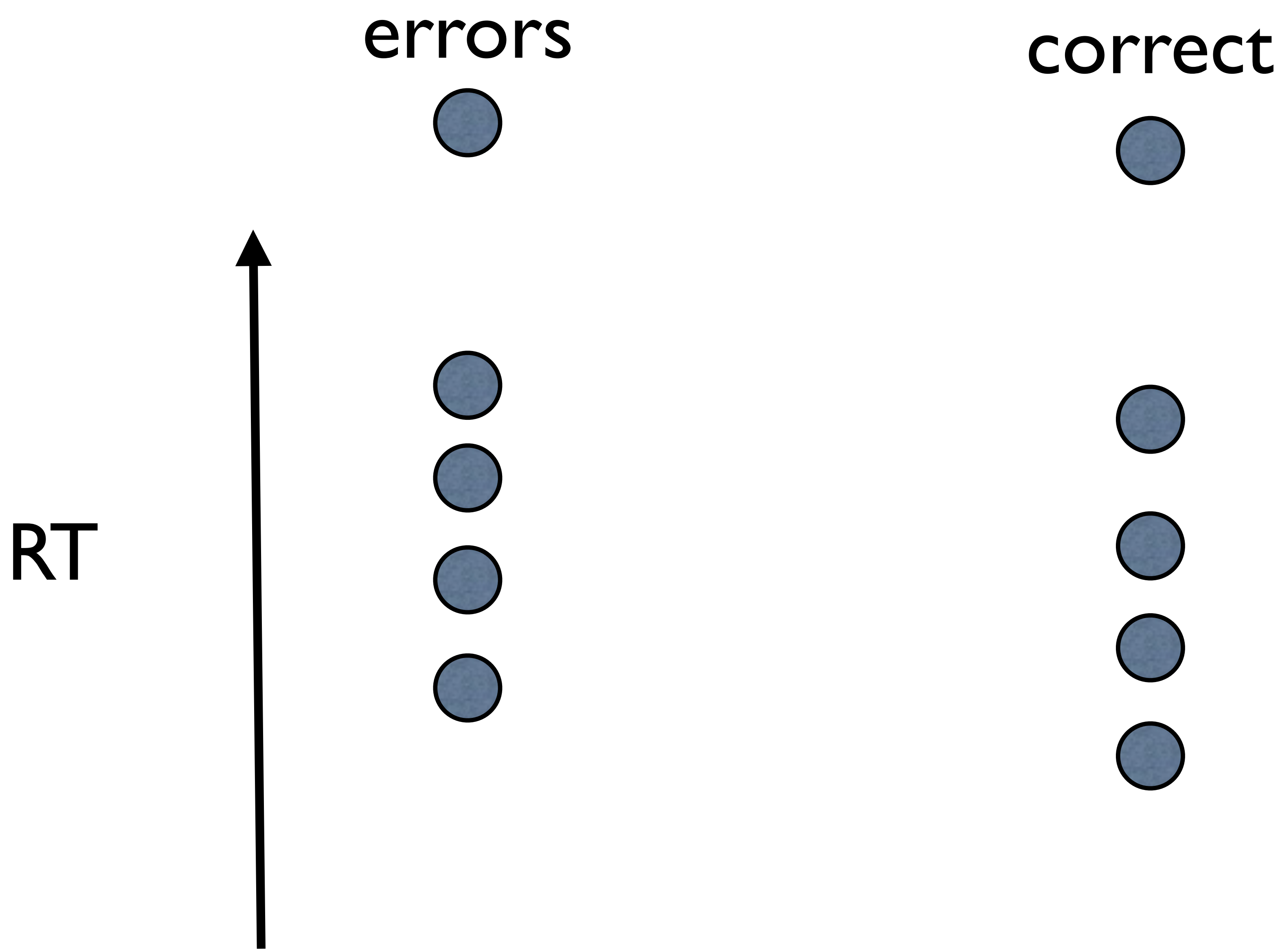
Plot the bin edges on a vertical scale

(note, normally we don't plot the rotated histogram, but it helps in this case to see what you're looking at)

Quantile probability functions

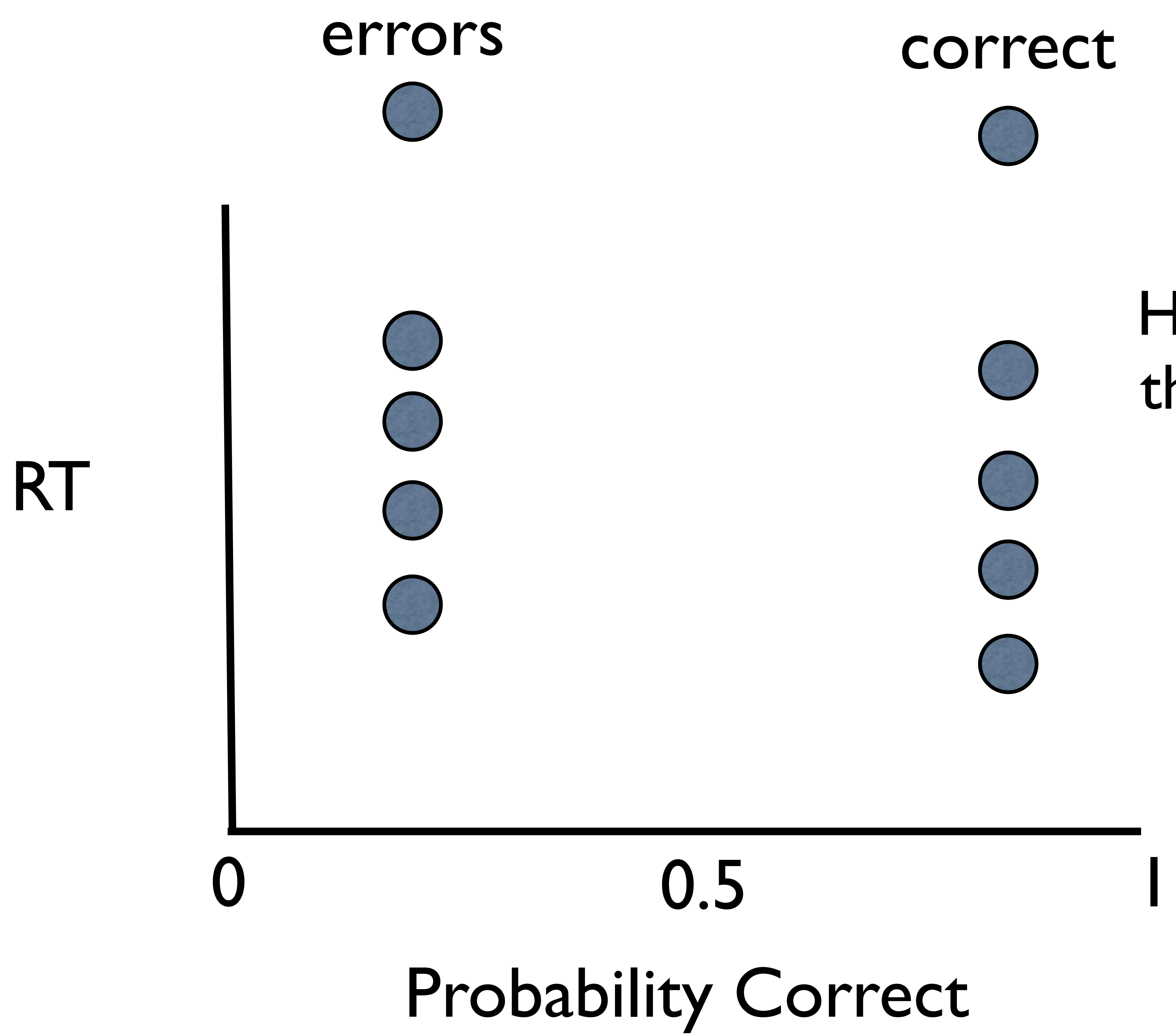


Quantile probability functions

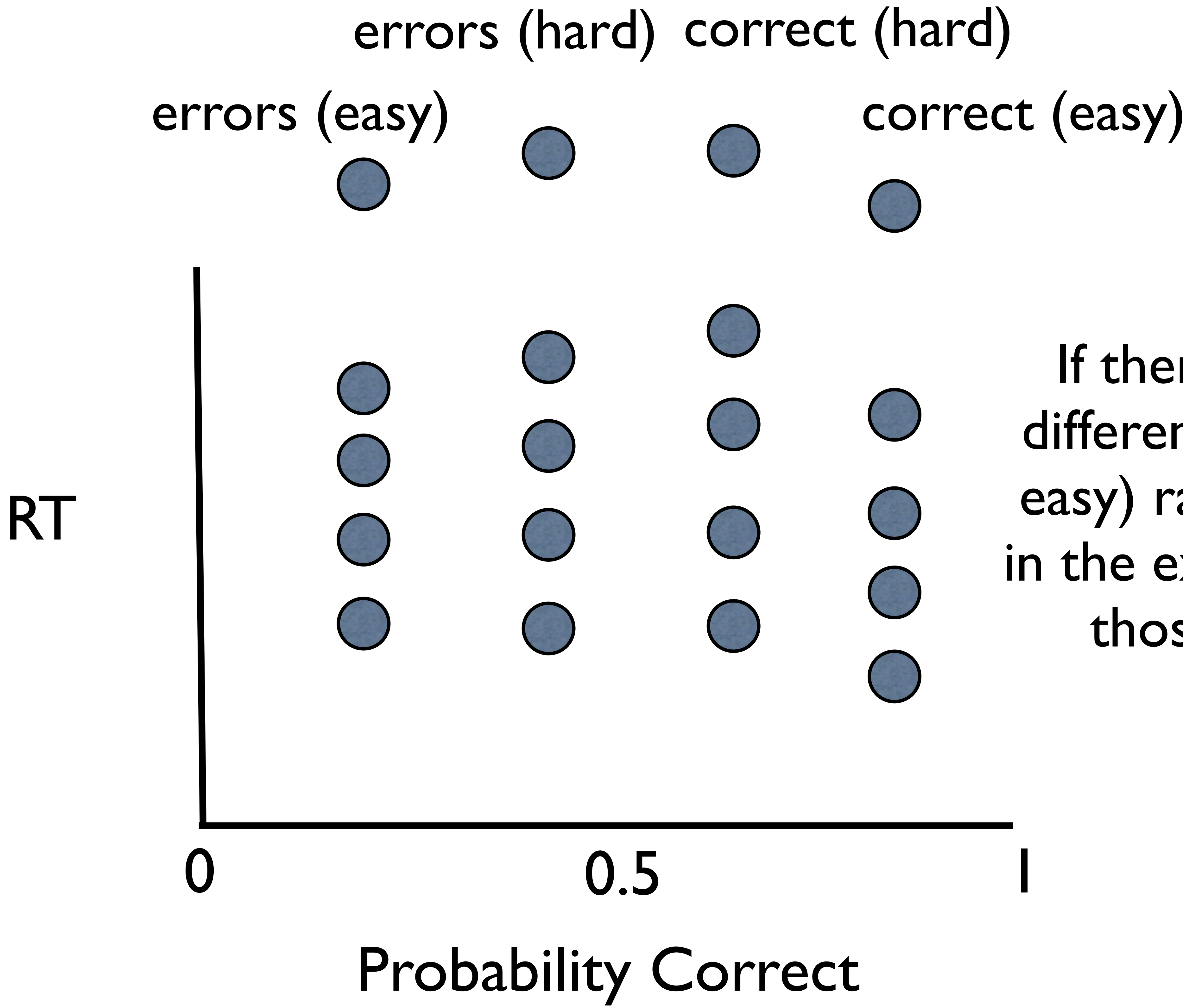


Plot "error" trials separately from "correct" trials, because error RT distributions are systematically different to correct RT distributions

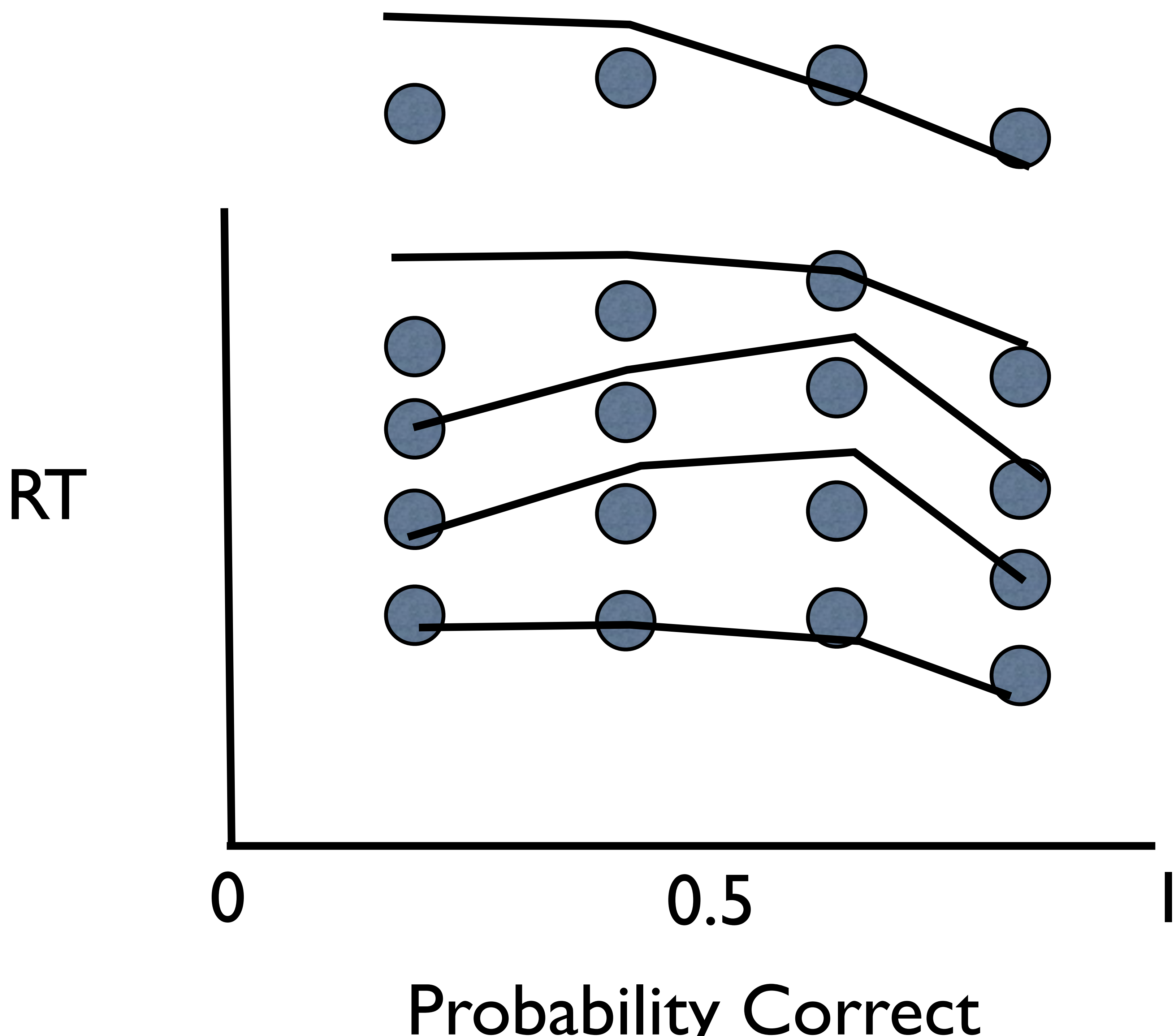
Quantile probability functions



Horizontal location reflects the probability correct (i.e., P(corr) for the correct trials, 1-P(corr) for the error trials)



If there are qualitatively different trials (e.g. hard vs easy) randomly intermixed in the experiment, then plot those separately too

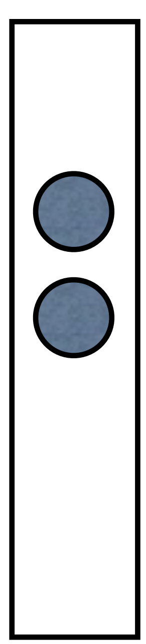


After fitting the model to the data, draw the model "predictions" as lines overplotting the data

Cognitive science application #1:
<u>Perceptual</u> decision making

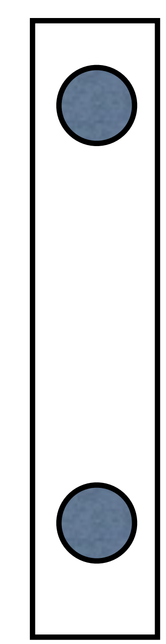
"Perceptual signal detection task"

Show people two vertically separated dots, ask them to classify as "small" or "large" separation



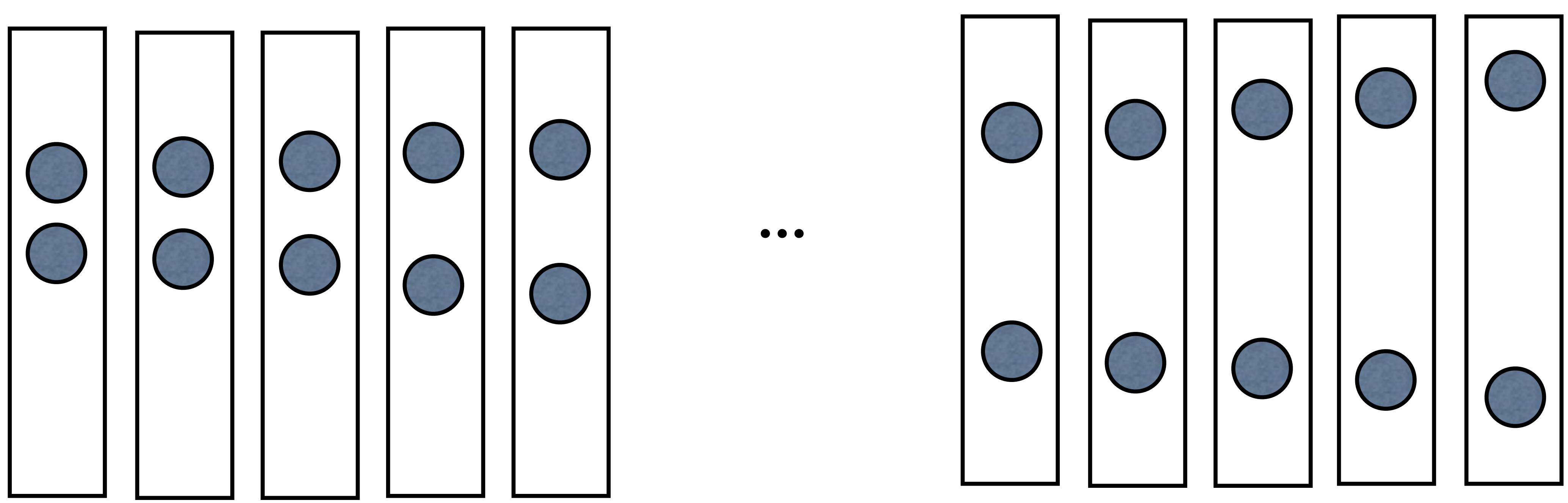
small

large

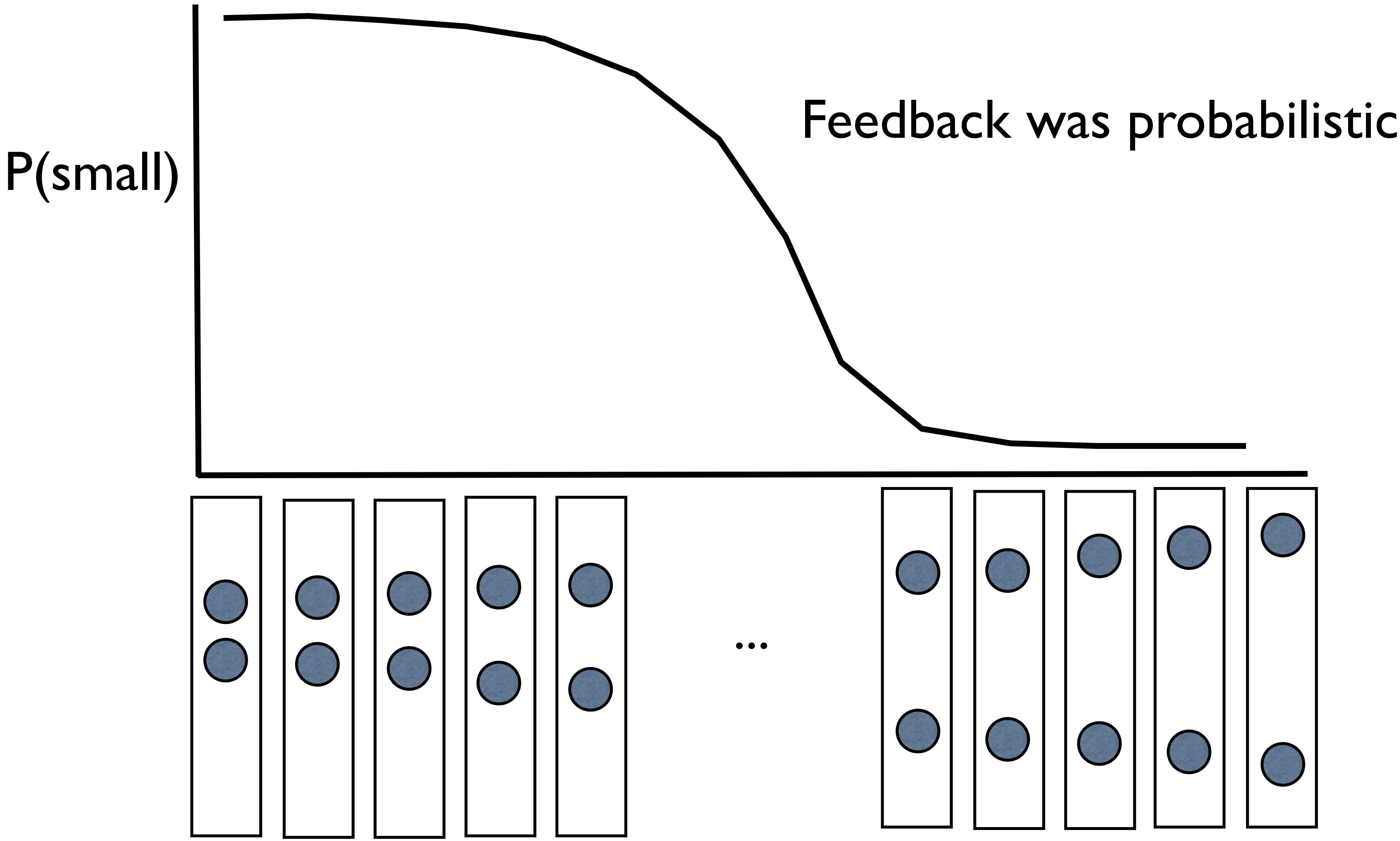


"Perceptual signal detection task"

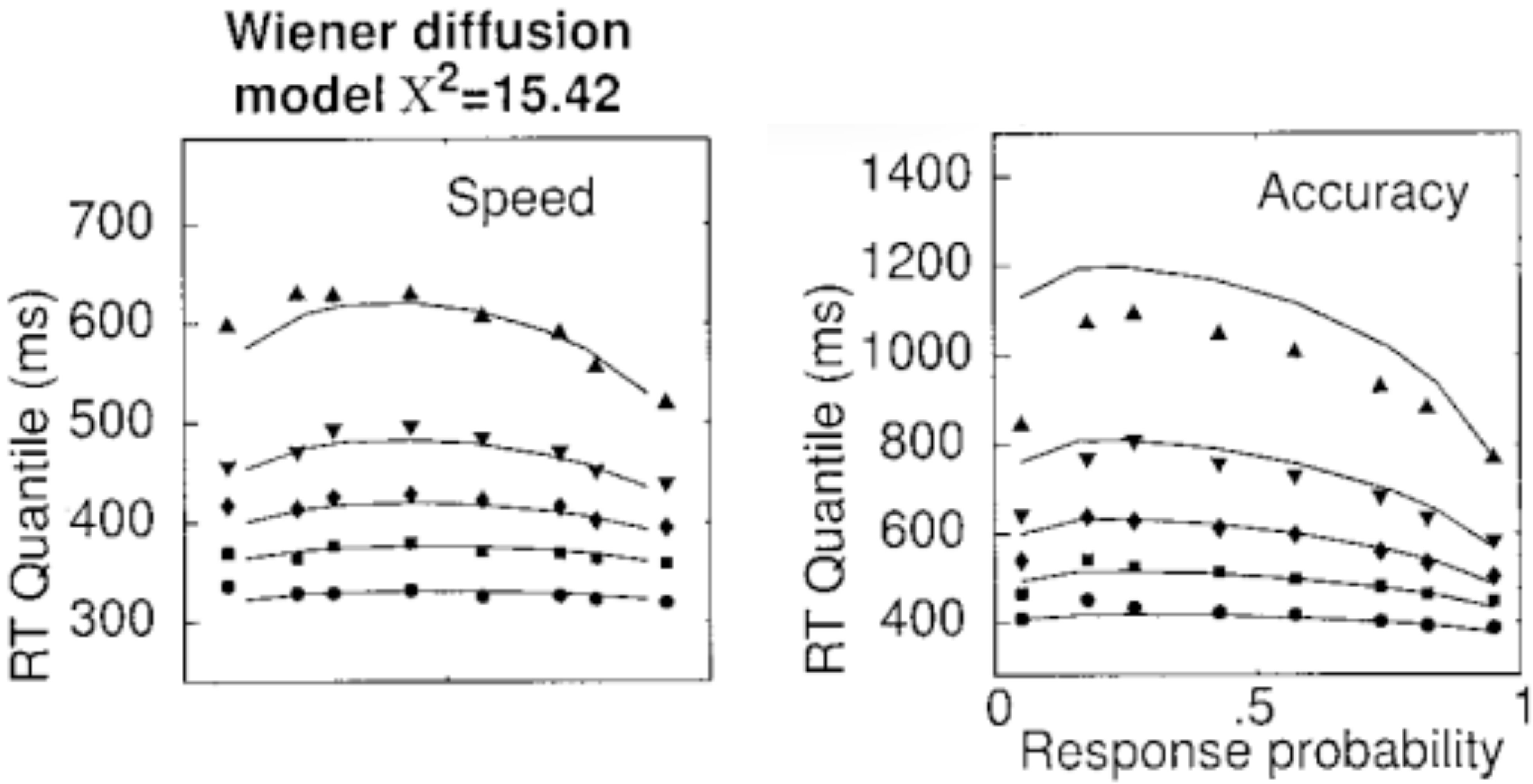
32 different "separations", repeated a very large number of times in randomised order. Participants were given feedback (i.e. told the right answer afterwards)



"Perceptual signal detection task"



"Perceptual signal detection task"



Model fits for two conditions: when "speed" is emphasised in the instructions to participants, and when "accuracy" is emphasised

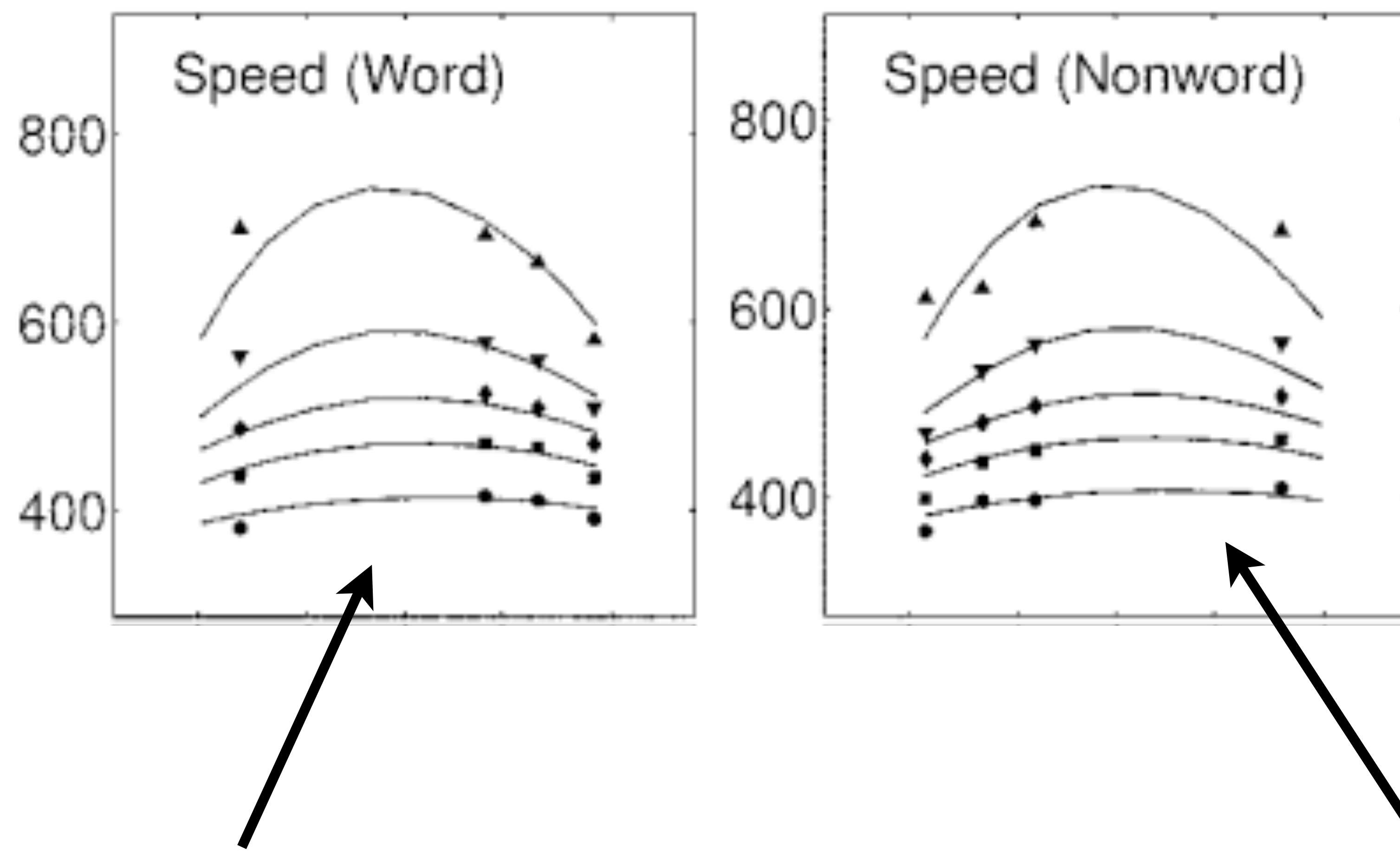
Cognitive science application #2: Lexical decision making

"Lexical decision task"

Task is to decide if the stimulus is a word or a non-word. Four types of trials:

Non-words:	Low freq. words:	Medium freq. words:	High freq. words:
Farbic	Enigmatic	Flicker	Music

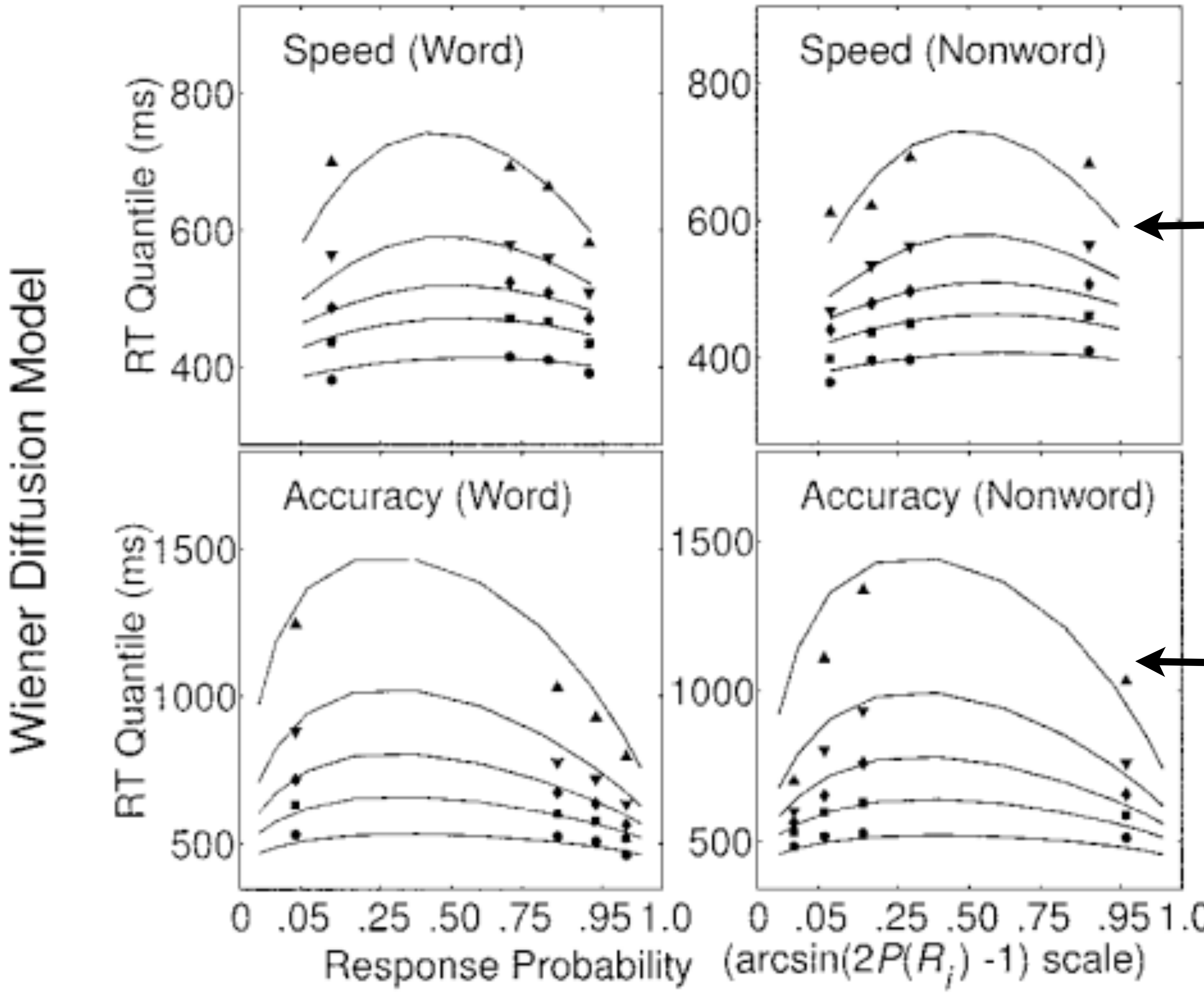
"Lexical decision task"



Trials when the person said "word" (horizontal location reflects the probability that people said "word")

Same thing, but for trials when people said "nonword"

"Lexical decision task"



Speed condition

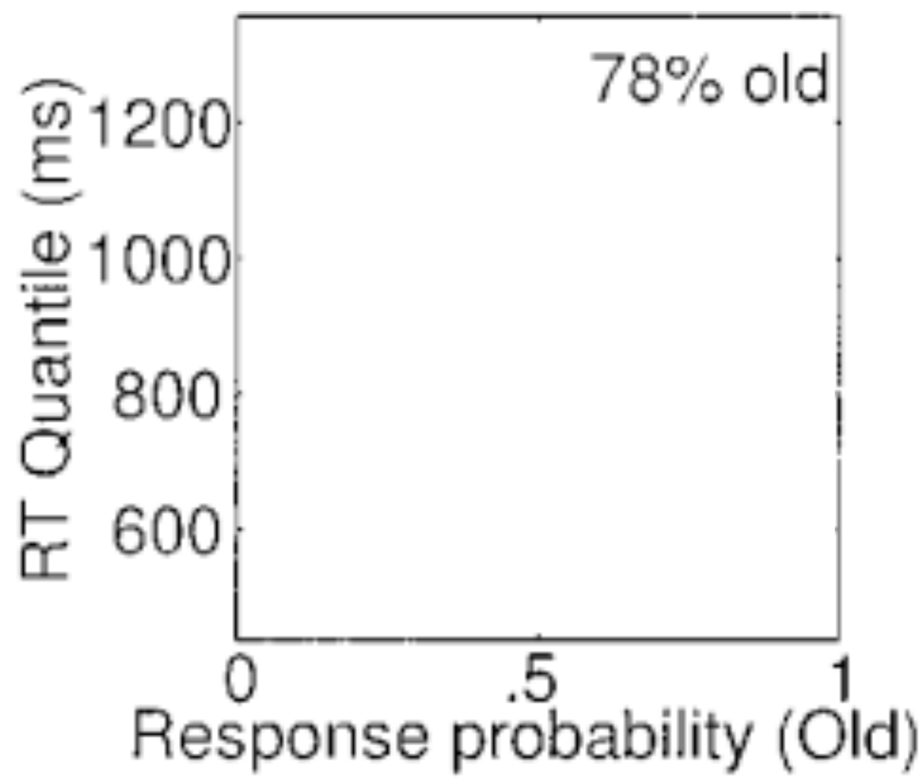
Accuracy condition

Cognitive science application #3:
Making decisions about your memories

"Recognition memory task"

- "Study list" of words to memorise (e.g., warm, tire, polearm, etc...)
- Later, show "test" words, and ask people to judge if it was in the study list (e.g., happiness, warm, stochastic, polearm, etc...)
  - Both the old & new items could either be high, medium or low frequency words
  - Different experimental conditions involve different proportions of old items and new items

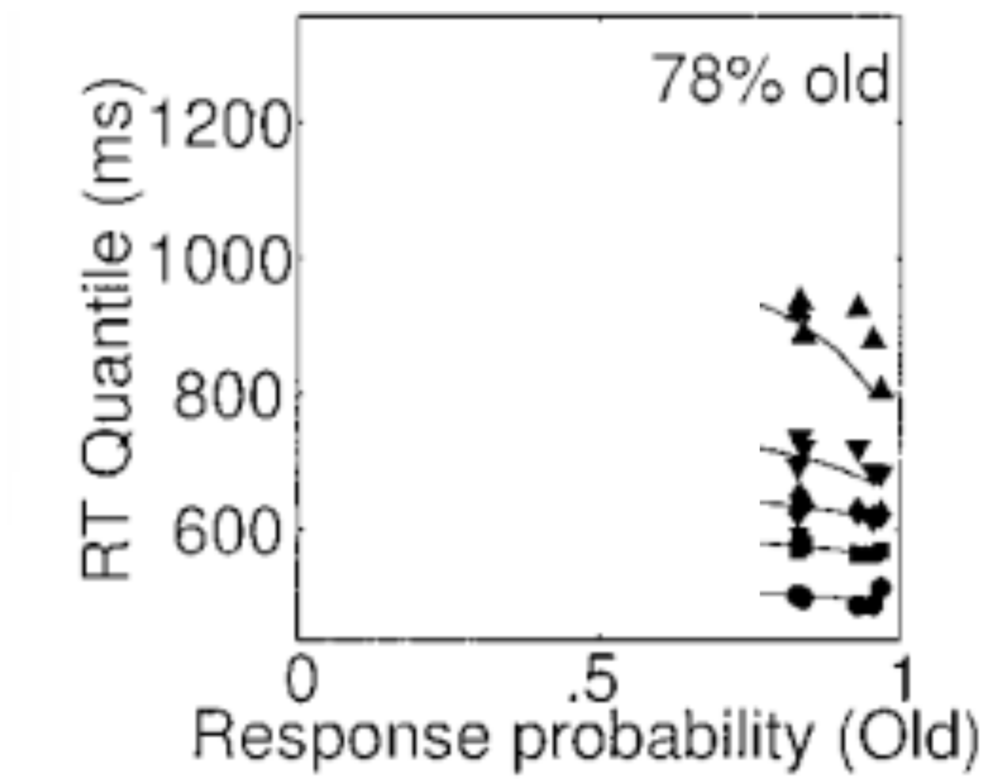
"Recognition memory task"



Trials where people said it was an "old item"

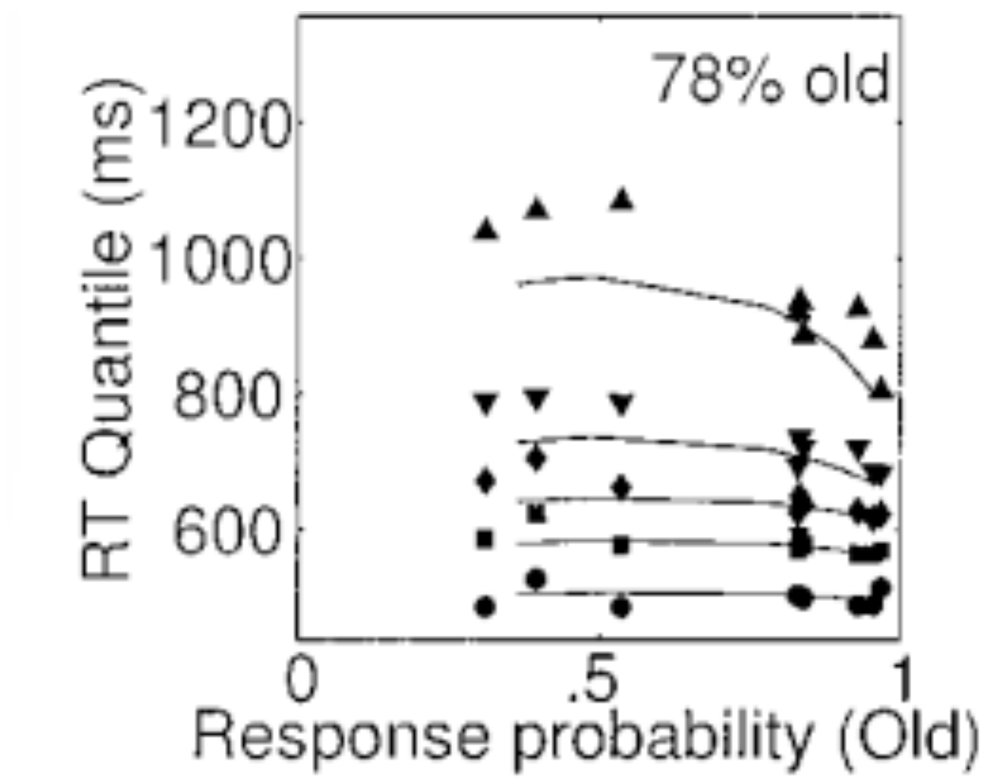
In an experimental condition when 78% of the items were actually old

"Recognition memory task"



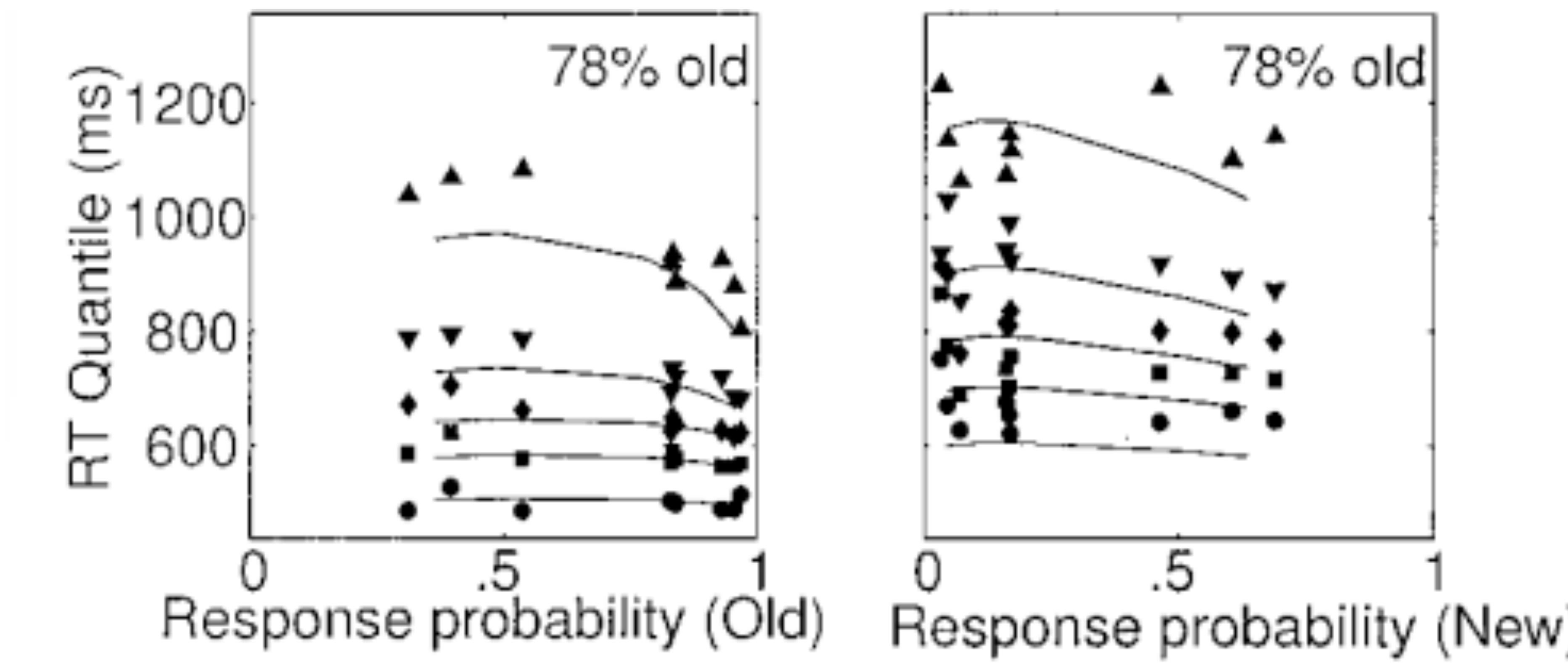
These are the trials when they were actually correct in saying "old", broken down by "high", "medium" and "low" frequency

"Recognition memory task"



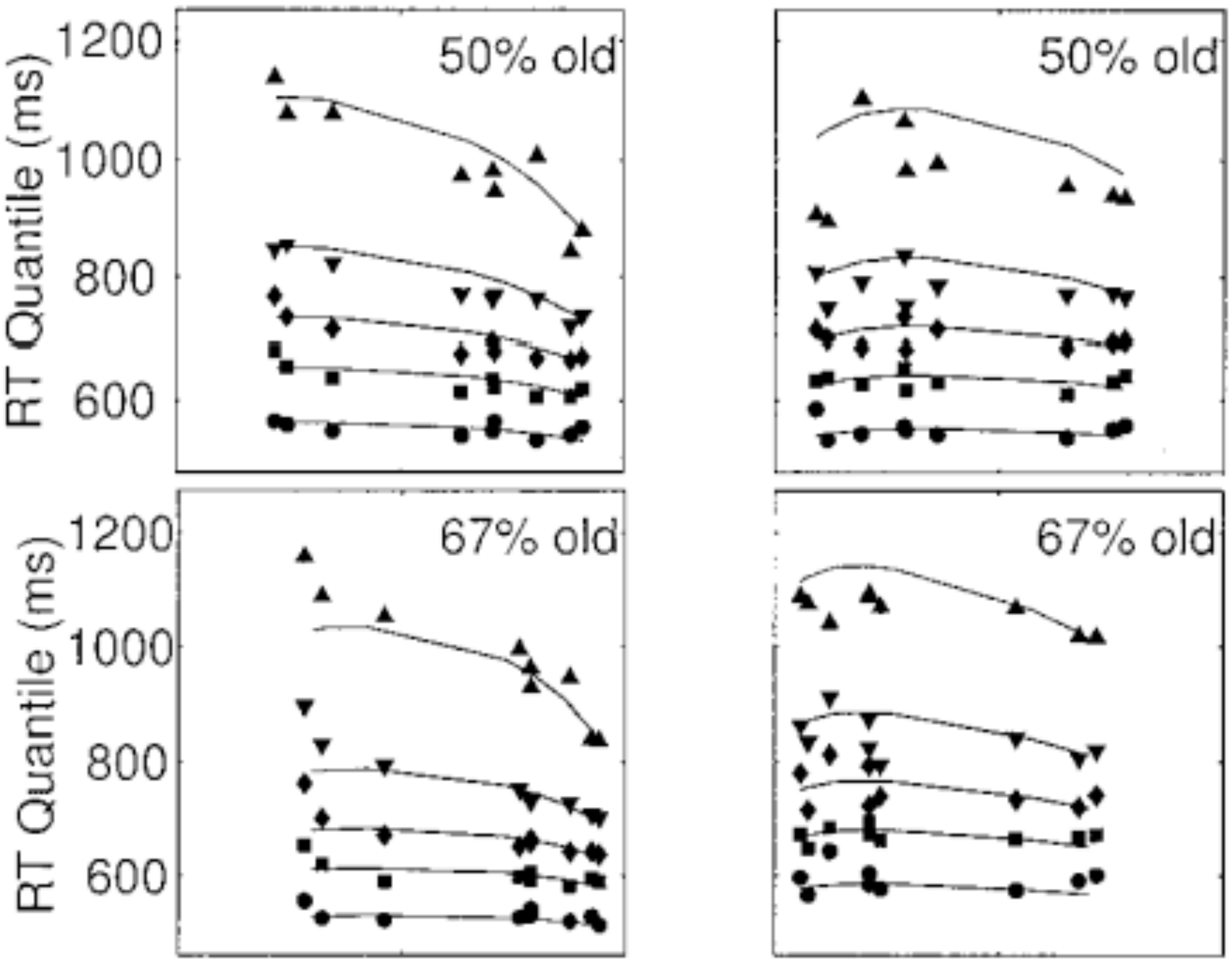
Similarly, trials where people were wrong in saying old

"Recognition memory task"



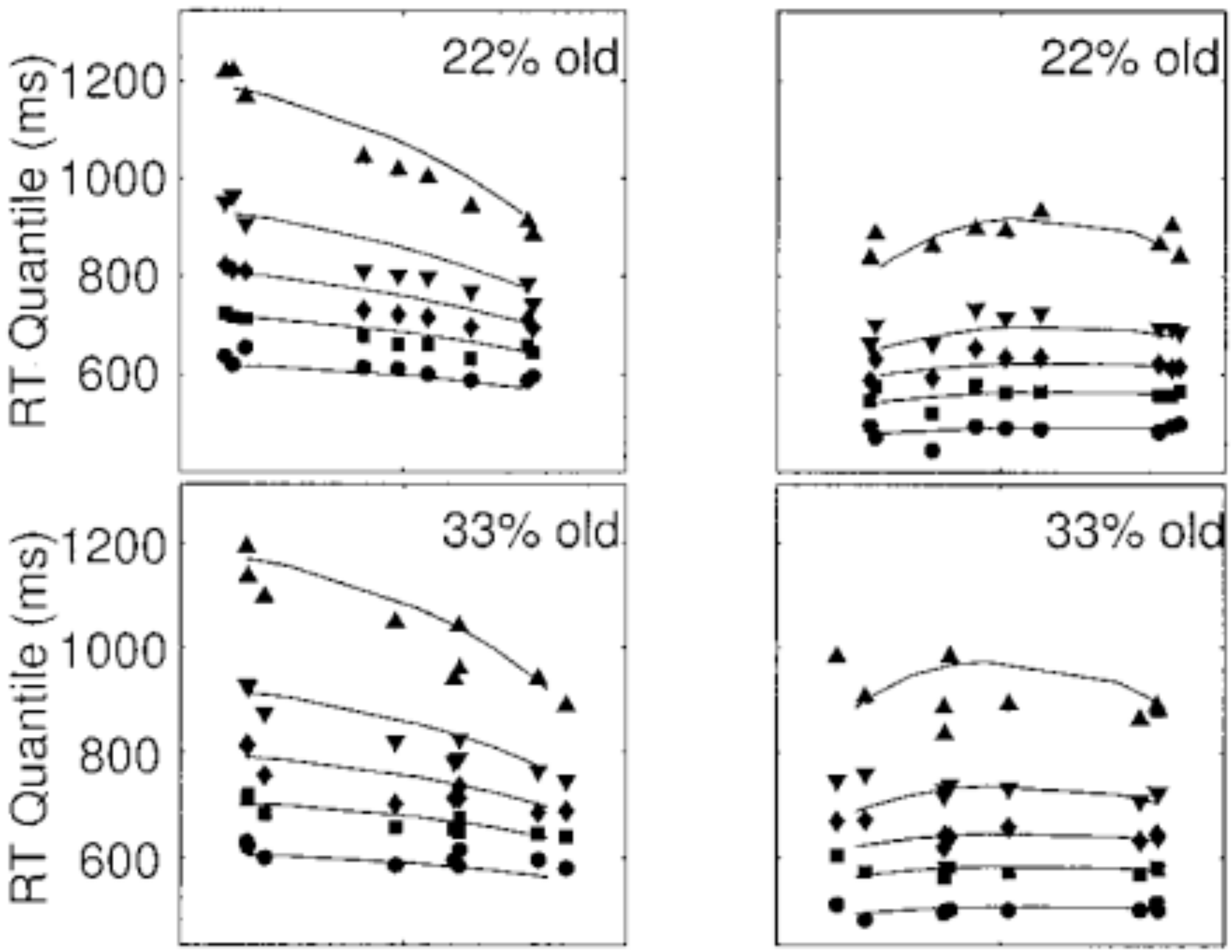
Same plot, but for trials
when people said "new"

"Recognition memory task"



Two more experimental conditions

"Recognition memory task"



And two more

Application to a machine learning problem:
Quick and not-so-dirty text classification

Is this a news article about SHIPPING?

Is this a news article about SHIPPING?

Wireless

Is this a news article about SHIPPING?

Wireless broadband

Is this a news article about SHIPPING?

Wireless broadband use

Is this a news article about SHIPPING?

Wireless broadband use has

Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed

Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed but

Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed but South

Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed but South Australia

Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed but South Australia remains

Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed but South Australia remains behind

Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed but South Australia remains behind the

Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed but South Australia remains behind the pack

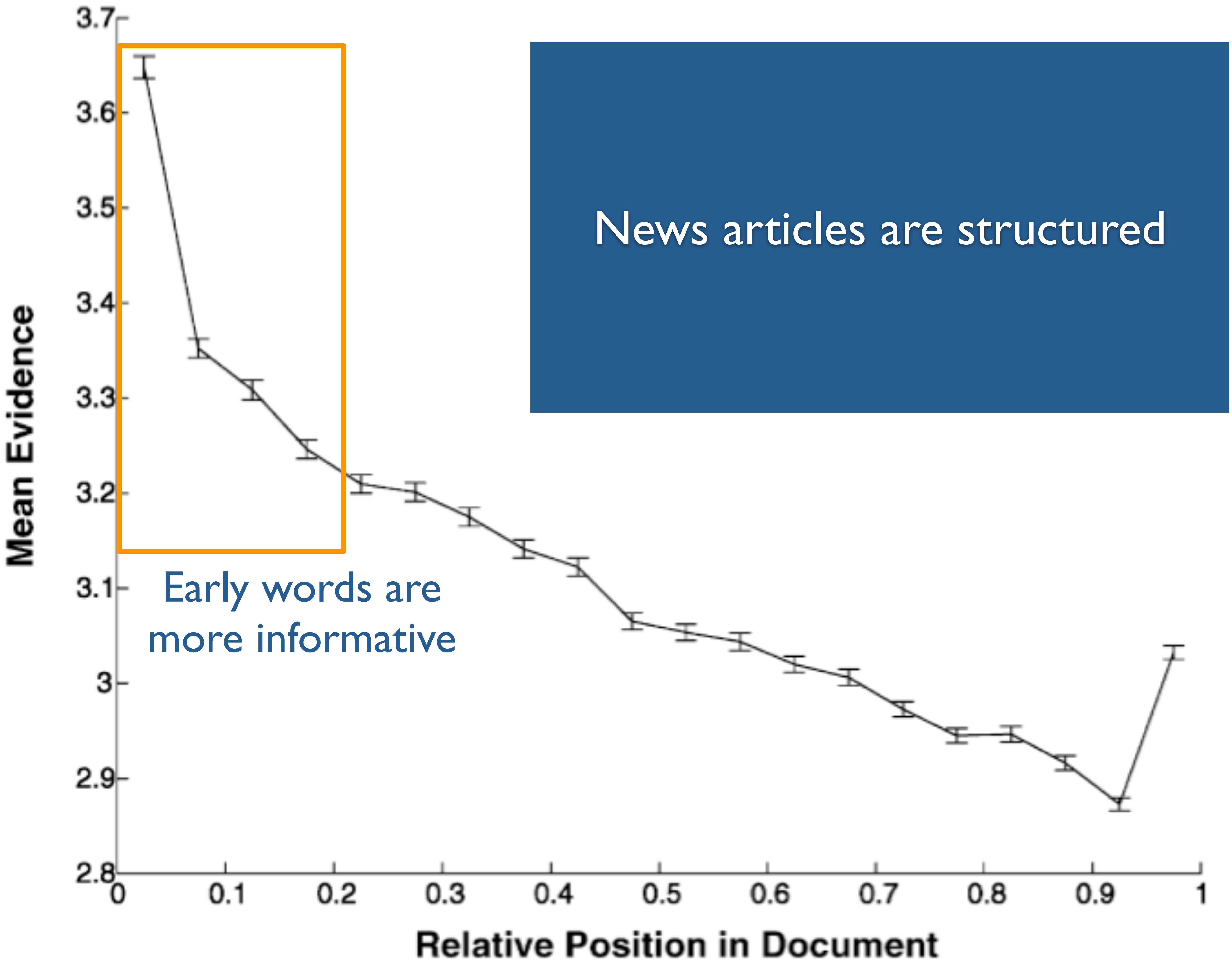
Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed, but South Australia remains behind the pack when it comes to access. Nationally, use of 3G mobile and wireless broadband services grew by 162 per cent during 2008-09 to reach 2.1 million services by June 30, the Australian Communications and Media Authority says. (Adelaide Now, Jan 14 2010)

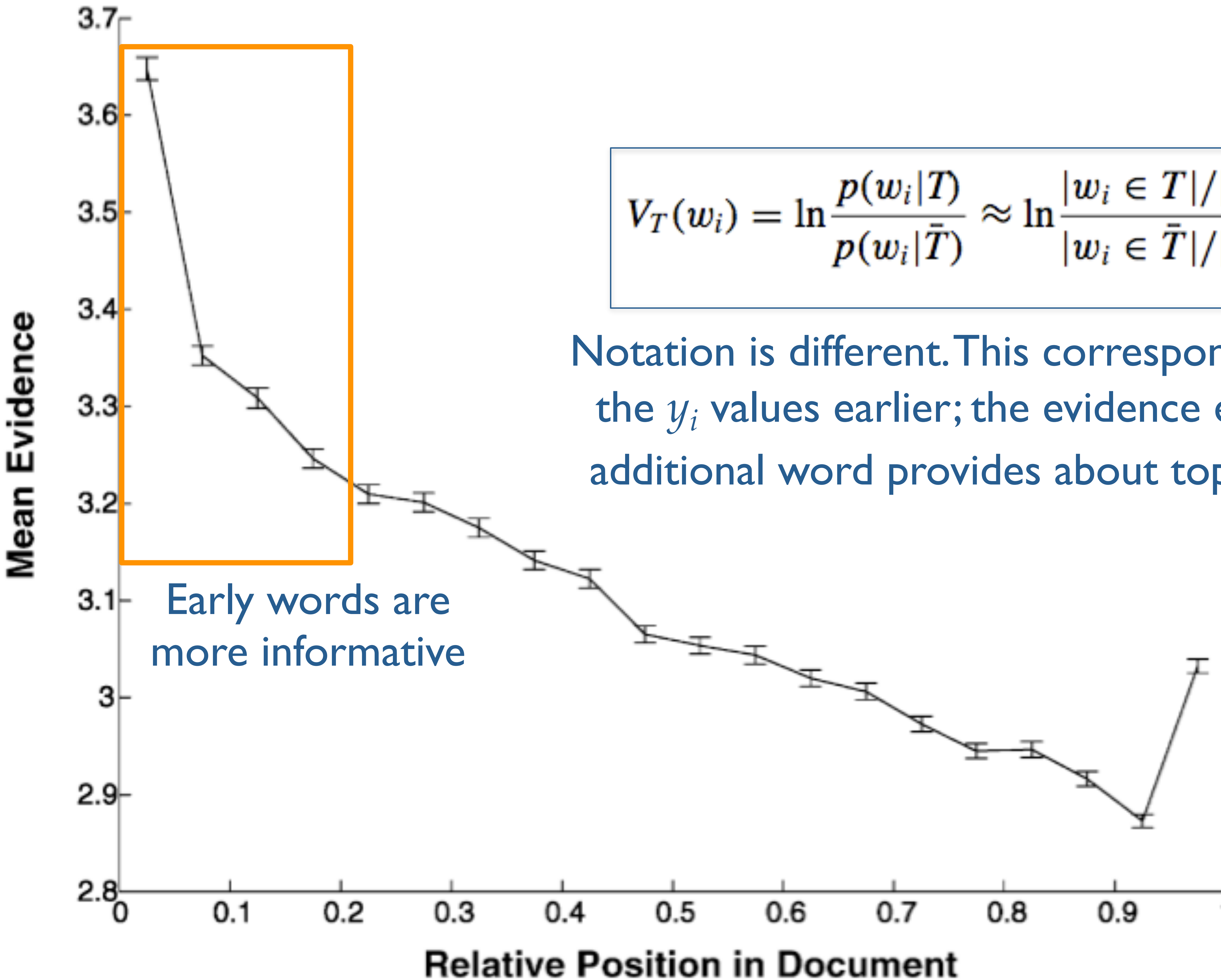
Is this a news article about SHIPPING?

Wireless broadband use has skyrocketed, but South Australia remains behind the pack when it comes to access. Nationally, use of 3G mobile and wireless broadband services grew by 162 per cent during 2008-09 to reach 2.1
Australian Communic
(Adelaide Now, Jan 14

You don't need to read the whole thing to figure out what it's about



News articles are structured
Early words are more informative
Mean Evidence
Relative Position in Document
3.7
3.6
3.5
3.4
3.3
3.2
3.1
3
2.9
2.8
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1



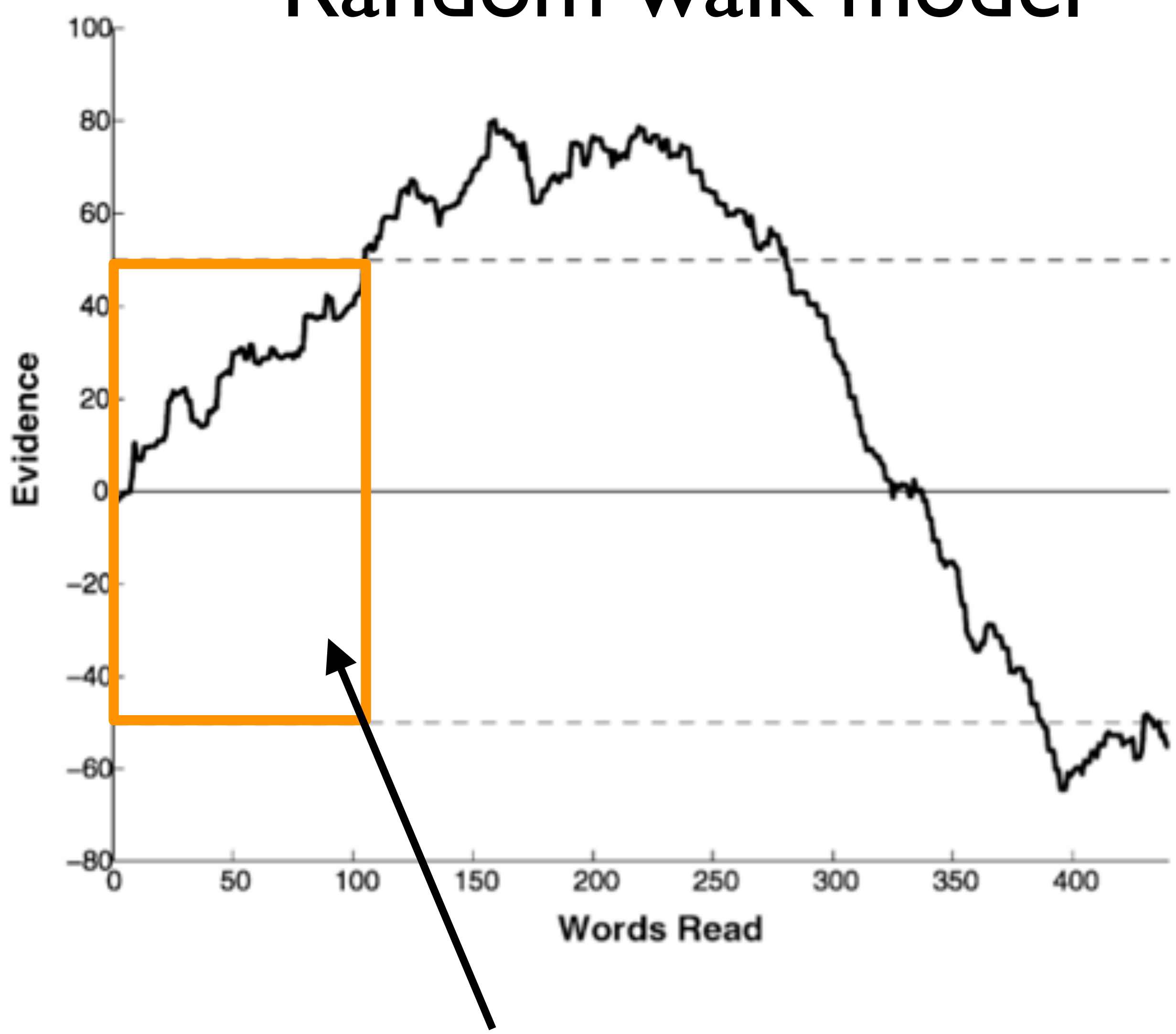
$$V_T(w_i) = \ln \frac{p(w_i|T)}{p(w_i|\bar{T})} \approx \ln \frac{|w_i \in T|/|T|}{|w_i \in \bar{T}|/|\bar{T}|}$$

Notation is different. This corresponds to the y_i values earlier; the evidence each additional word provides about topic T

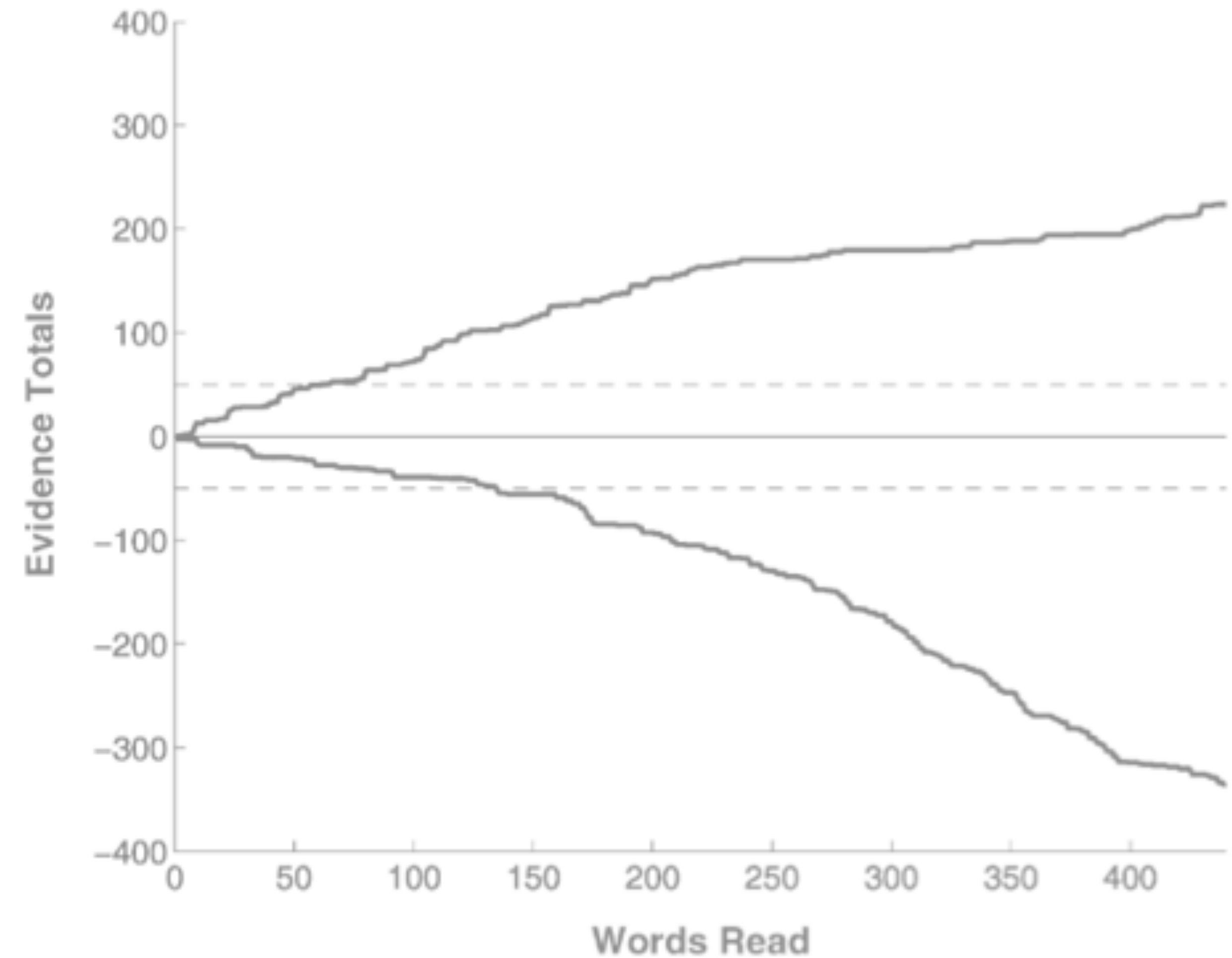
Early words are more informative

Two sequential sampling models

Random walk model

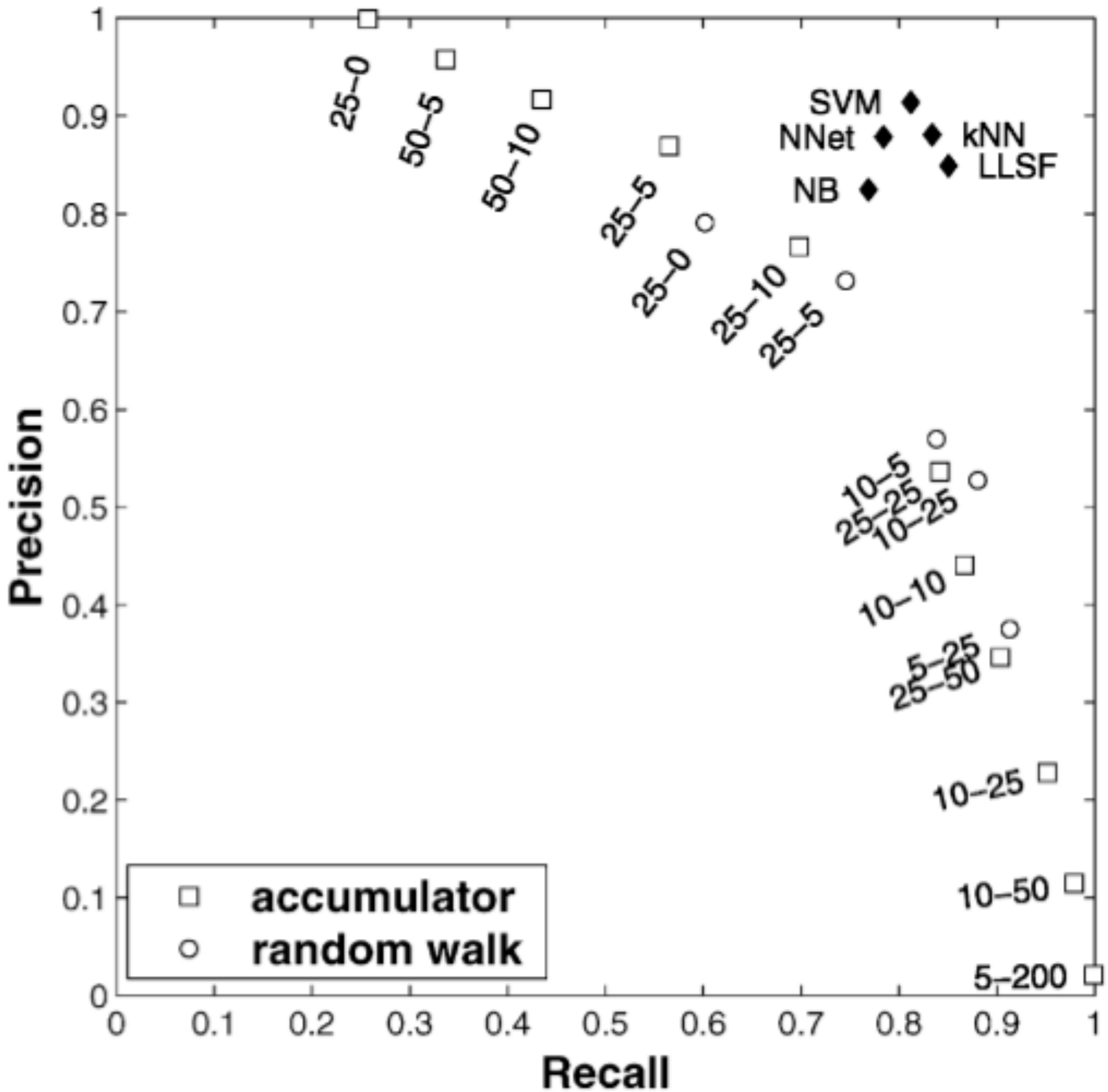


Accumulator model



"Read" the words, one at a time, incrementing the evidence counter, until a decision boundary is reached: In this case, stop after about 110 words, and decide "yes, this document is about topic T"

Precision-recall curves



Performance close to best known machine learning methods, but requires NO pre-processing of the text corpus, scales linearly with the number of documents, and typically only read 5-6 words in each document regardless of length!

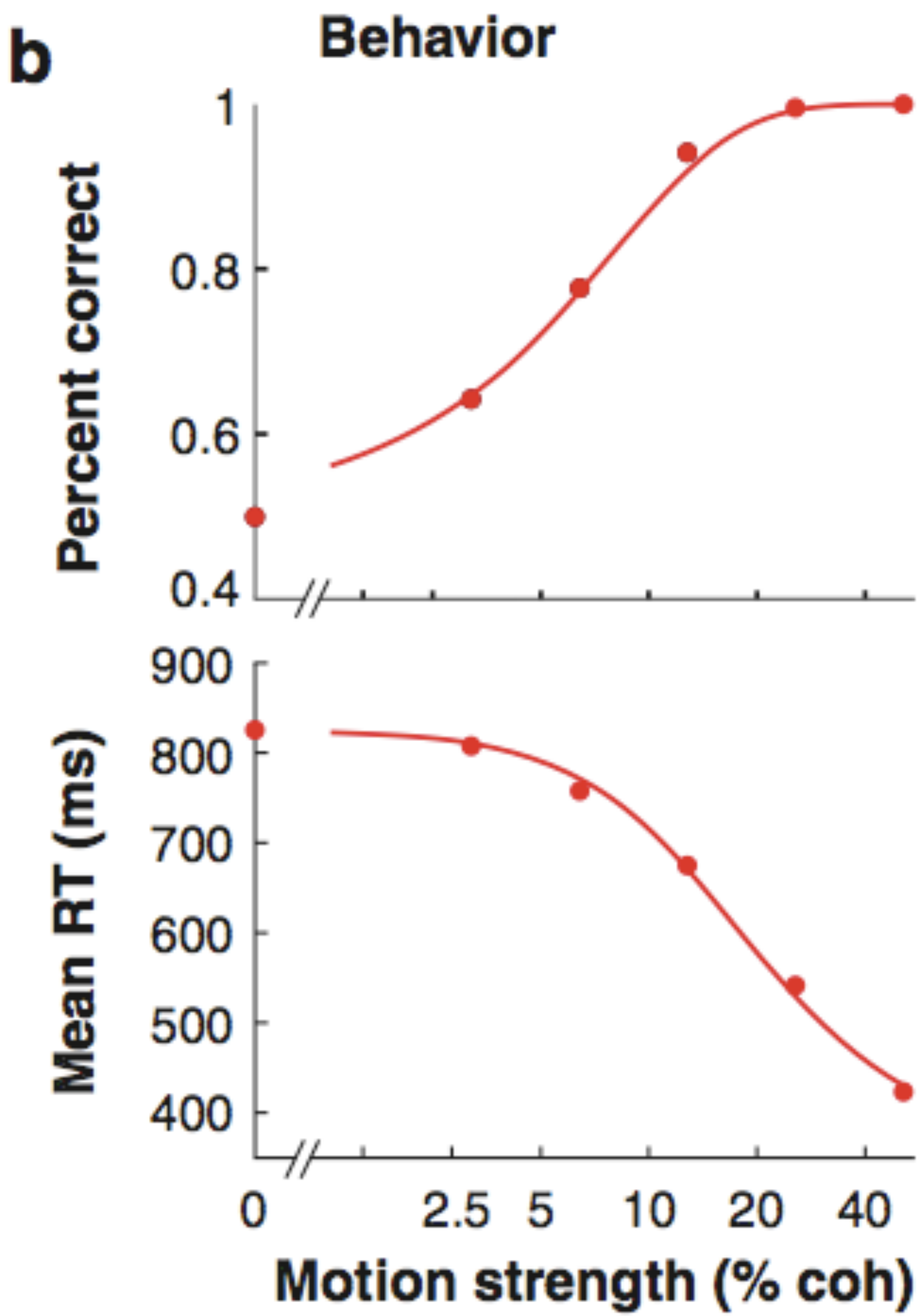
By comparison SVM requires all documents to be processed in full, and requires a quadratic programming problem to be solved

Application to neuroscience: Evidence accumulation in monkey brains (no, really!)

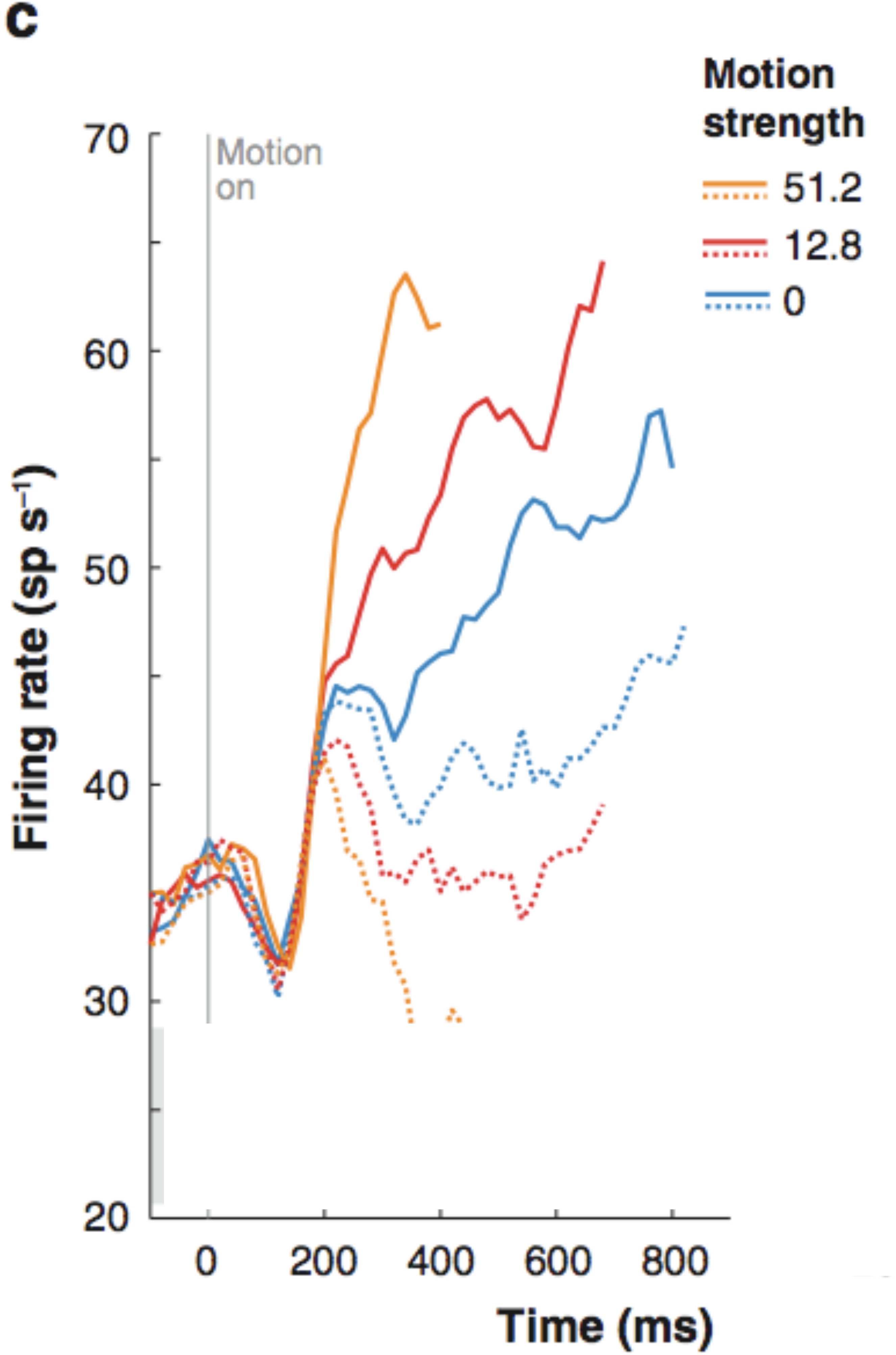
From psychology to neuroscience

- The "evidence tallies" in SSMs are theoretical ideas.
  - Not directly observable.
  - At least, not until recently.

- In recent years, neuroscientists have started using the same modelling tools, and have been able to find neural systems that behave like diffusion models!

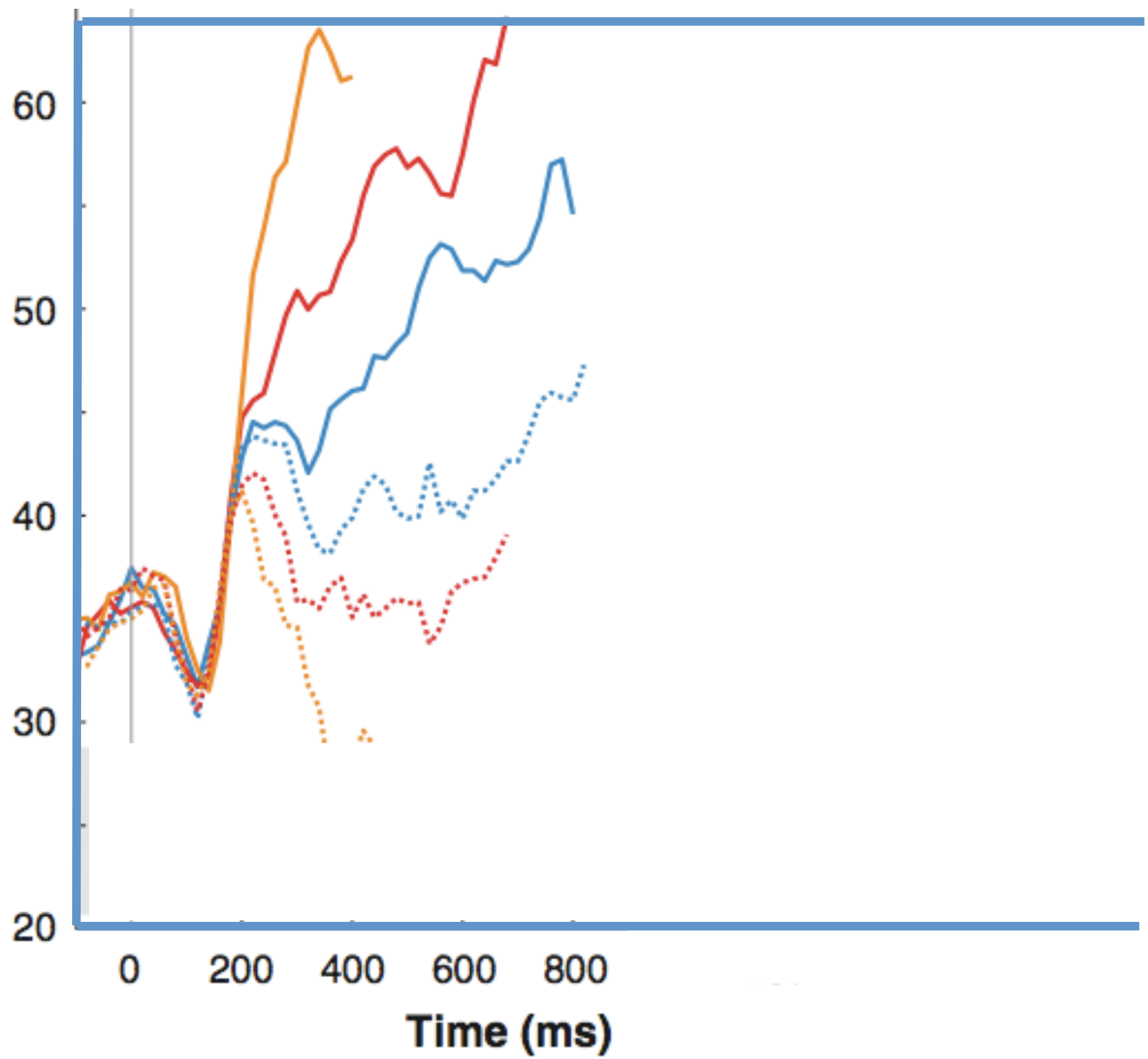


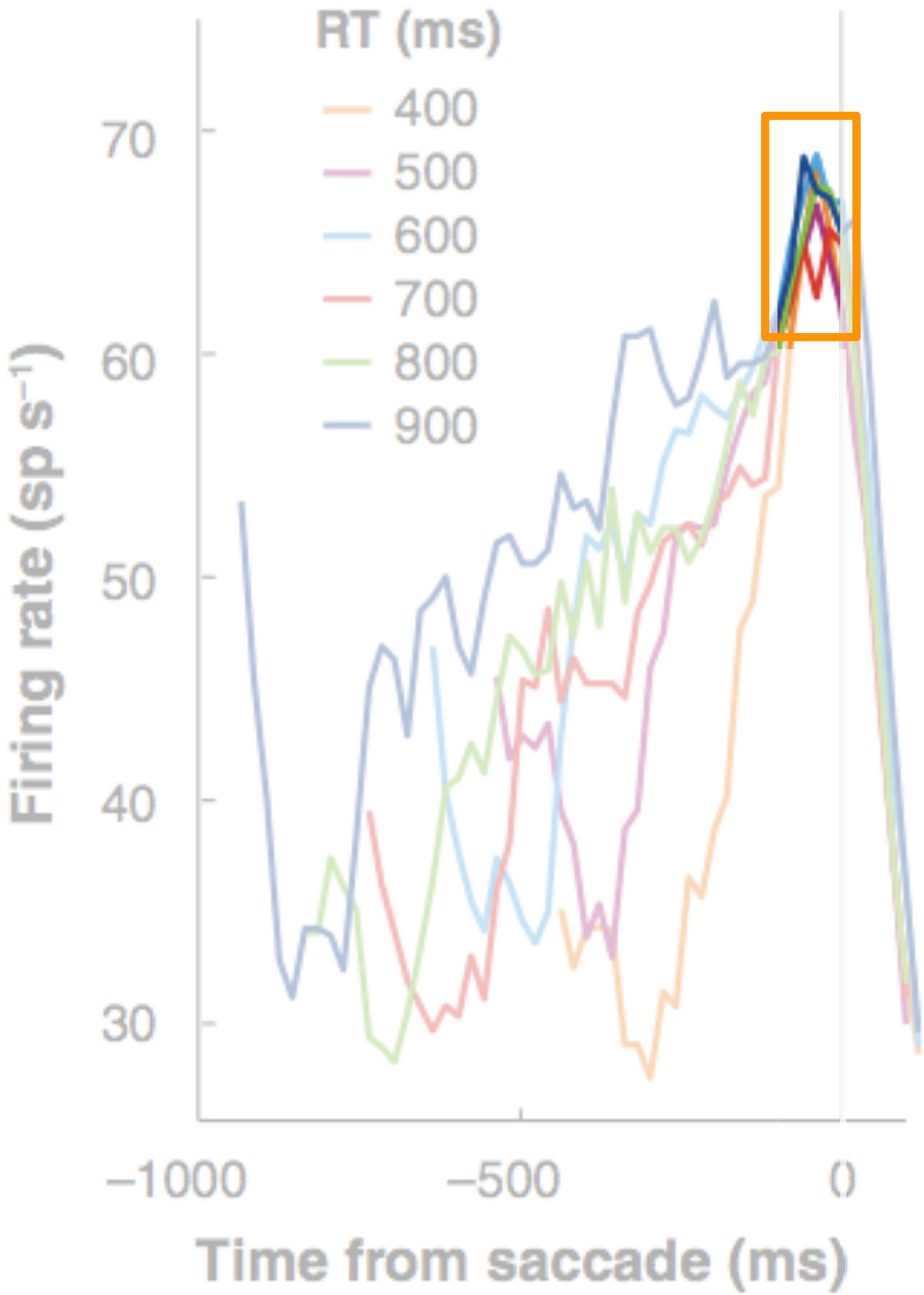
Standard psychometric functions... as the decision becomes easier, monkeys (like humans) become faster (bottom) and more accurate (top)



Firing rate of selected neurons (neuroanatomical details omitted!) as a function of time, broken down by motion strength (i.e., ease of decision)

Looks suspiciously like a sequential sampling model?





The neural firing data, looking "backwards in time" from the moment of the saccade, as a function of RT...

Looks very much like the decision is initiated at a fixed "firing rate" level?

i.e., fixed decision threshold.

Further reading

Selected References

- Sequential sampling models generally
  - Ratcliff, R. & Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. Psychological Review, 111, 333-367.

- Quick calculations for the diffusion model
  - Navarro, D. J. & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. Journal of Mathematical Psychology, 53(4), 222-230
  - See the RWiener package in R

Selected References

- The text classification example
  - Lee, M.D., & Corlett, E.Y. (2003). Sequential sampling models of human text classification. Cognitive Science, 27(2), 159-193

- The neuroscience side (not discussed in the lecture)
  - Gold, J.I. and Shadlen M.N. (2007) The neural basis of decision making. Annual Review of Neuroscience, 30, 535-574.

Selected References

- If you've got an interest in the mathematics behind sequential sampling models (i.e., first passage times for stochastic processes):
  - Smith, P. L. (2000). Stochastic, dynamic models of response times and accuracy: A foundational primer. Journal of Mathematical Psychology, 44, 408-463.
  - Note: this paper is hard.