Drift-diffusion models

Mehdi Senoussi

Université de Toulouse Jean-Jaurès – CNRS

Laboratoire CLLE

mehdi.senoussi@univ-tlse2.fr
mehdisenoussi.github.io/
@cogsenoussi.bsky.social

Plan

- Why decision-making is cool (coming from someone who considered it as ***uninteresting*** for a long time ☺)
- Why modelling it? (I mean, we have accuracy and RT, what else do we need to characterize how people decide?)
- How can we model it/what do each family of models offer?
- The Drift-Diffusion Model (DDM): history, basic principles, and… ***fitting contest!***
- Getting to know the tools we'll use in this session
  - Python
  - Google colab (or local, e.g. Vscode, Spyder, Pycharm, etc.)
- The proper pipeline for a study using the DDM (using pyddm on colab)
  - [*long list, we'll see that later*]
- Cool existing/possible additions to the core DDM
- Examples of cool results thanks to the DDM

Importance of decision-making

- **Addiction**: individuals make decisions to sacrifice long-term health for short term benefits of the addictive substance or behavior
- Economic behavior
- Neurodevelopmental diseases (e.g. ADHD)
- etc…

➢ Better understanding cognitive & neural mechanisms of DM is fundamental to improve psychological/pharmacological treatment (*and treatment adherence!*)

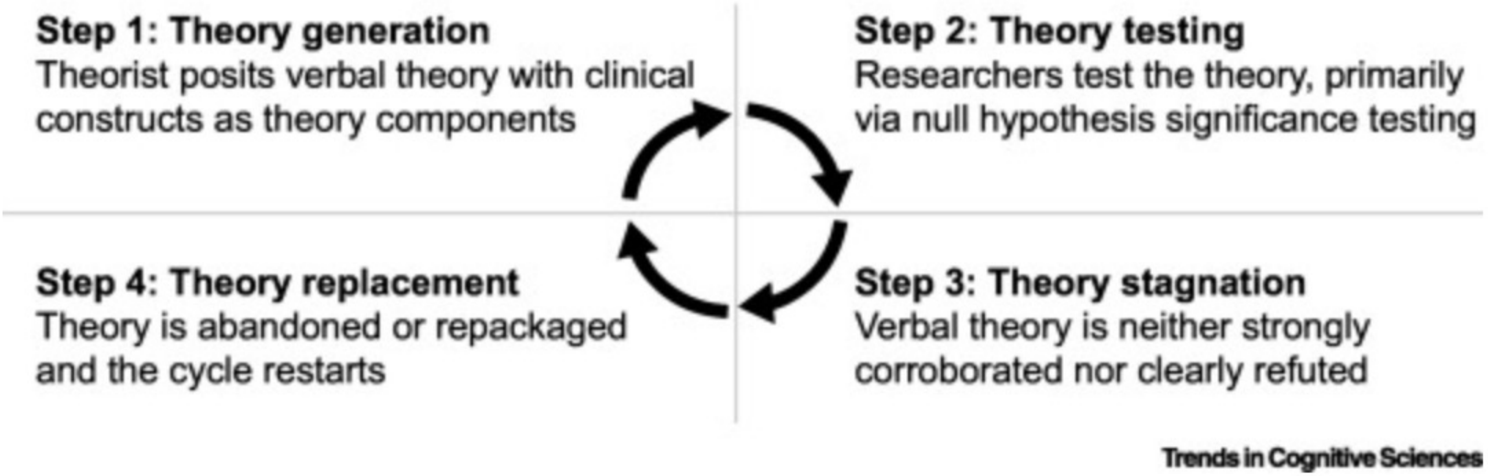
The *MAIN* goal: measure ***latent variables***

(to understand underlying mechanisms)

- The general idea behind the DDM (and most other computational approaches): measure hidden variables relevant to understand/predict behavior/neural activity
- We measure accuracy & RTs but
  1) They vary across trials even with exactly the same stimuli/contexts, we need to understand why
  2) We want to understand the information processing pipeline/mechanism, but a pattern of accuracies and RTs could be compatible with a range of mechanisms

Why modelling?

The example of theories of suicide:

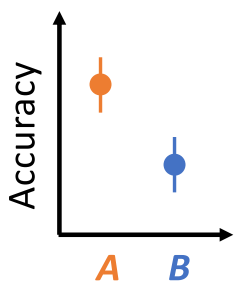


Millner et al. (2020)
Advancing the Understanding of Suicide: The Need for Formal Theory and Rigorous Descriptive Research

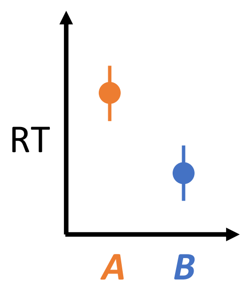
When a DDM is useful

- Speed-accuracy tradeoff (SAT) (partially under conscious control)
  - Diff pps have/use different SAT, complicating the interpretation of behavioral data

Hypothesis: *condition **A** is easier than **B***



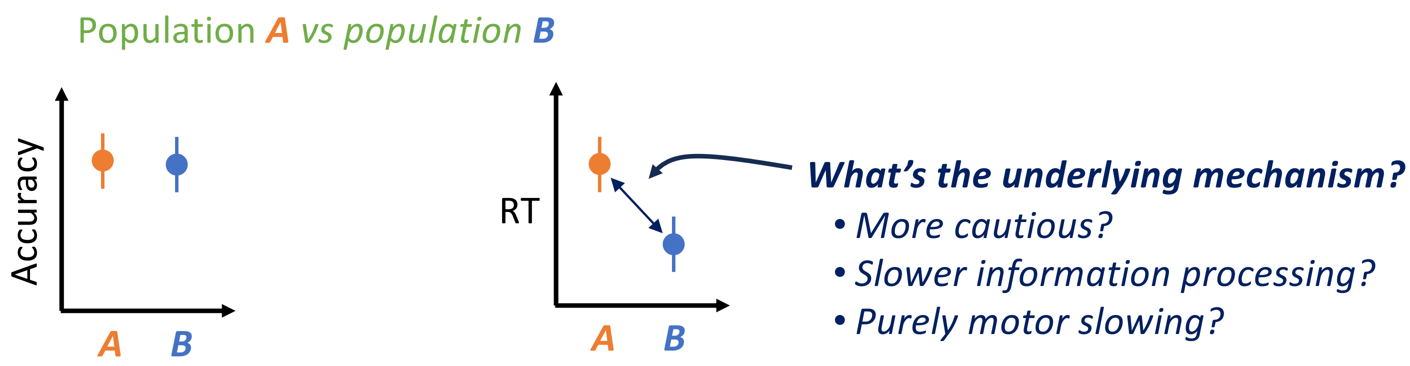
Wooooo YES!!



Oooooh… what??

When a DDM is useful

- Disentangling possible underlying causes for an effect



Need to consider the interaction between accuracy and RT to disambiguate possible explanations

A bit of history

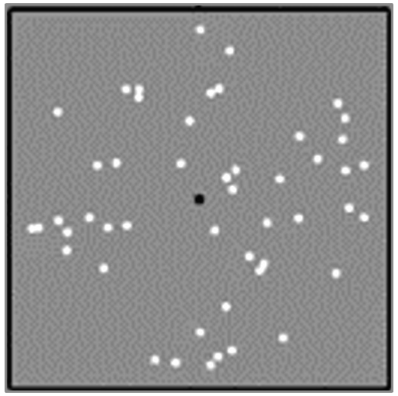
- Historically, the study of decision-making in psychology has been closely connected to the study of sensation and perception
- Two strands of this tradition are relevant:
  - Psychophysics: relationship between stimuli physical magnitudes and sensations (Gustav Fechner's « *just noticeable difference (JND)* » , ~1860)
  - The study of reaction time or response time (RT) by Franciscus Donders (~1868) who developed methods to measure the speed of mental processes (inspired by Helmholtz's work on nerve conduction speed)

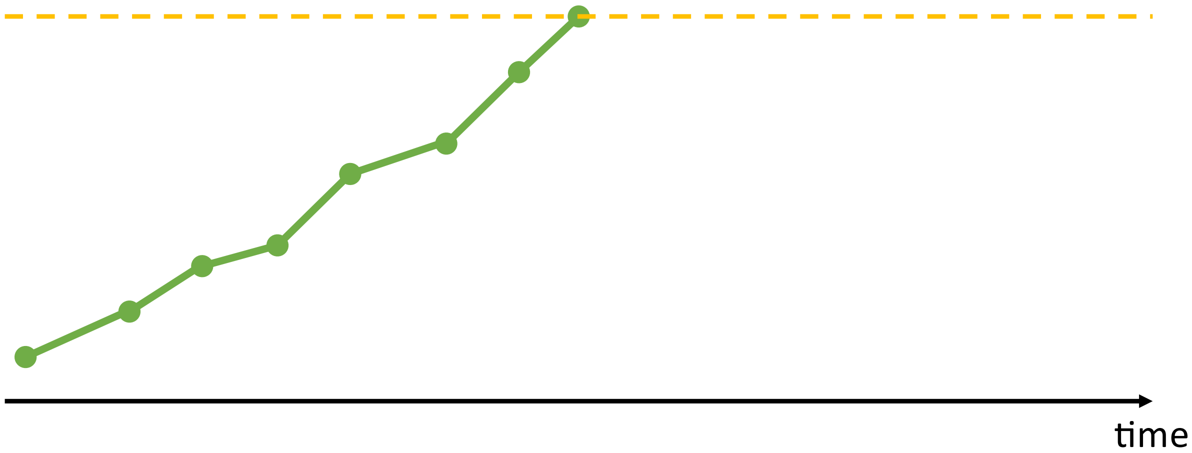
➢ Both strands realized that for the exact same stimuli/task/contexts/etc. the response and the time to produce this response varied significantly

➢ ***This led to adopt a view of decision as a noisy accumulation of evidence***

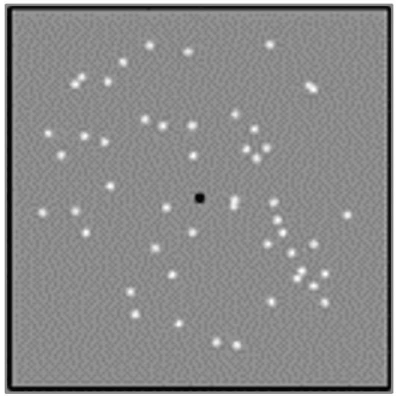
Smith & Ratcliff (2023)

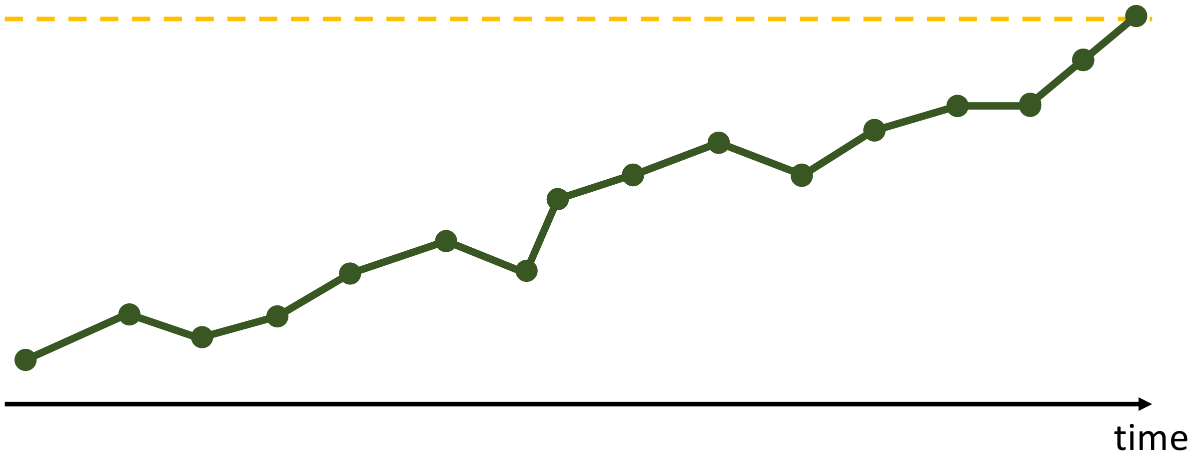
Basic assumptions: accumulating evidence



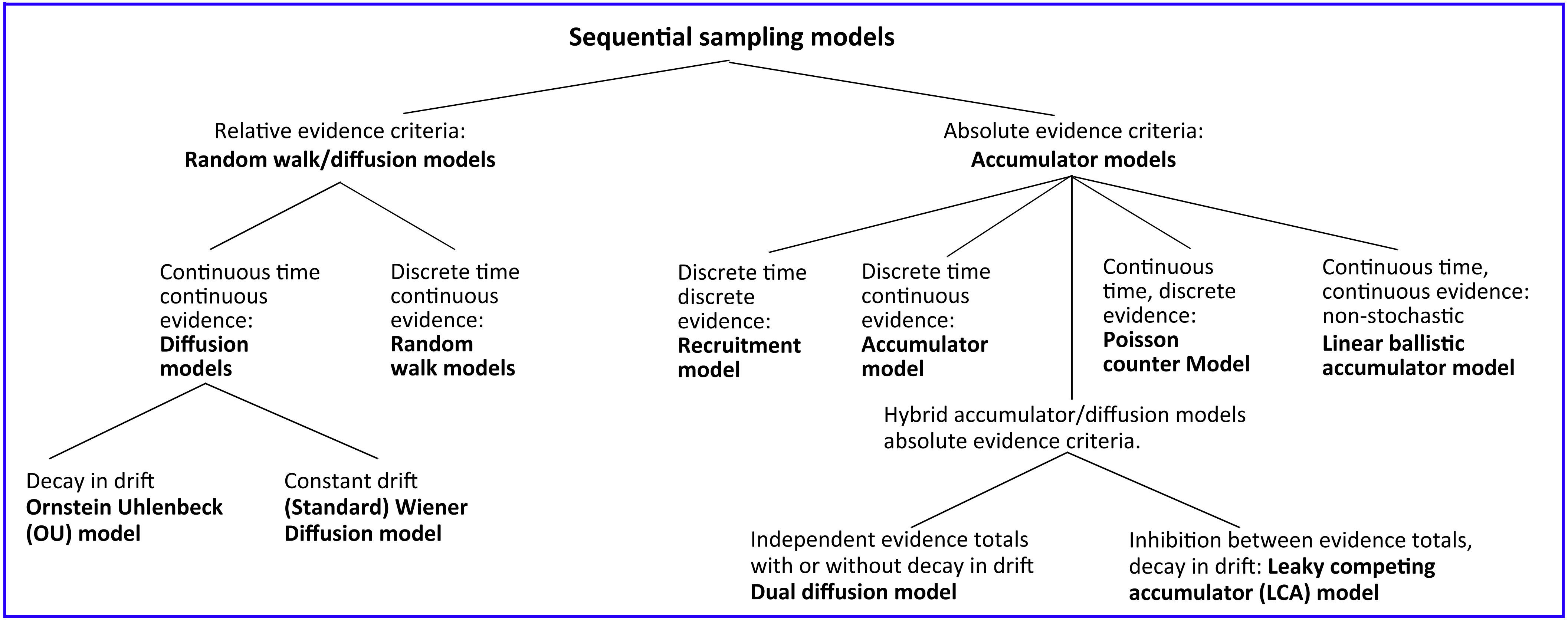


Basic assumptions: accumulating evidence





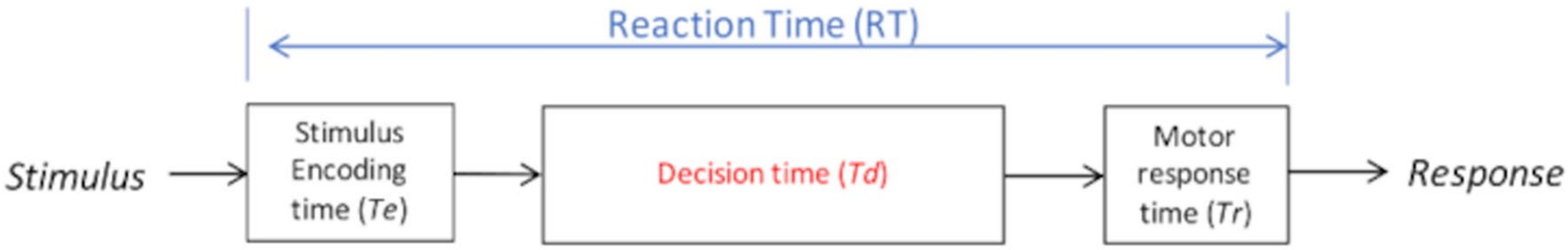
The buffet



Ratcliff et al. (2016)

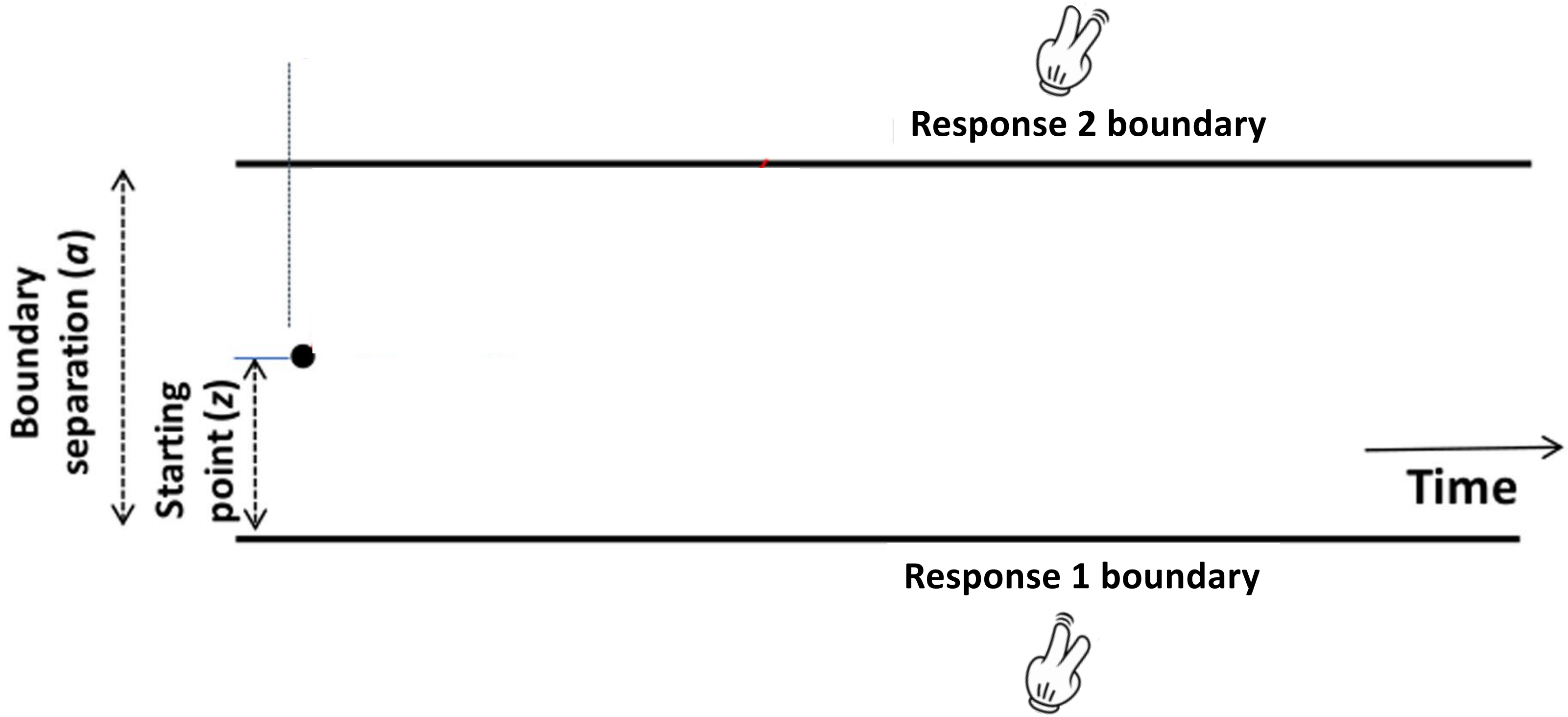
DDM assumptions

- RT = $T_{encoding}$ + $T_{decision}$ + $T_{response}$
  - ➢RT = $T_e + T_d + T_r$
  - T_e & T_r can in principle be measured separately but in the DDM they're lumped together in a T_{er} parameter representing « non-decision time »



Myers et al., 2022

DDM (hypothesized) process

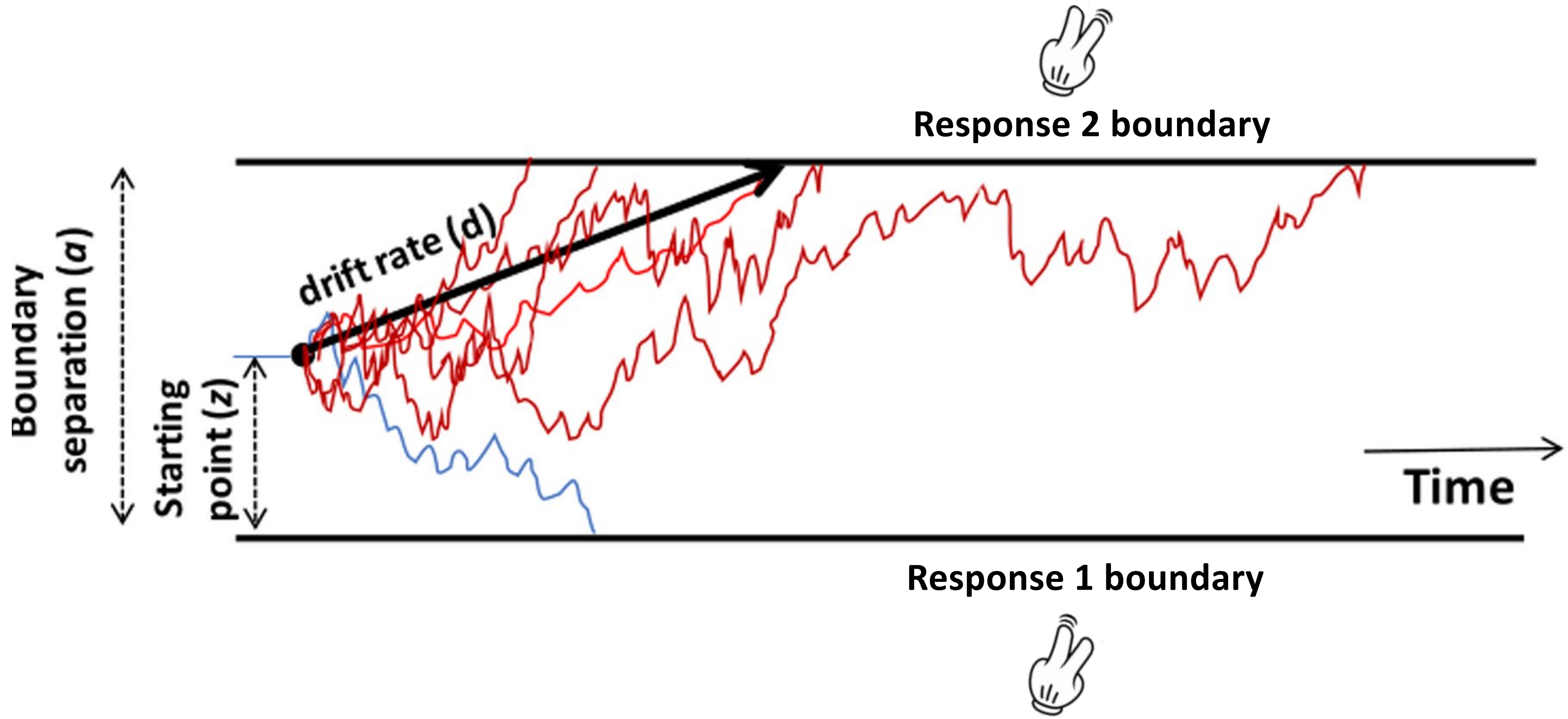


1 trial

Myers et al., 2022

DDM (hypothesized) process

$$dX_t = \xi dt + s\, dW_t.$$



multiple trials

Myers et al., 2022

DDM parameters

DDM parameter	Parameter name	Typical range of values	Cognitive processes
a	Boundary separation	0.5–2 (in arbitrary units)	Response caution: higher *a* emphasizes accuracy over speed, lower *a* emphasizes speed over accuracy.
z	Starting point	0…1 (as proportion of *a*)	Response bias: starting point nearer to one boundary leads to faster and more common decisions favoring that response.
d	Drift rate	−5…+5 (values <0 slope down to lower boundary)	Speed of evidence accumulation processing: can be affected by task difficulty, stimulus discriminability, attention.
Ter	Non-decision time	0.1–0.5 s (cannot exceed total RT)	Neurological processes for registering (encoding) sensory stimuli and for executing motor responses.

Myers et al., 2022

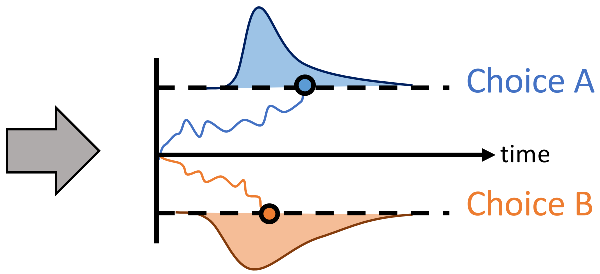
Some conceptual/terminology variability that can be confusing in DDM

- The boundary·ies/bound·s (*a*) (also called threshold/criterion sometimes)
  - Represents a threshold of accumulated evidence
  - More concretly: the « distance » separating thresholds corresponding to each choice
  - ***Intuitively,*** one could expect that: boundary choice 1 = $+\frac{a}{2}$ and choice 2 = $-\frac{a}{2}$
    - **BUT**, by convention (partly due to the original DDM article (Ratcliff, 1978)) lower boundary = 0 and upper boundary = *a* (in most packages)

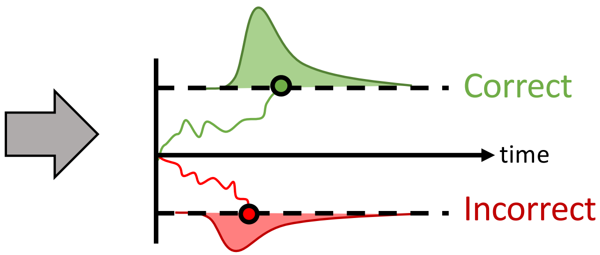
- An unbiased starting point is thus at $\mathrm{z} = \frac{a}{2}$ in most packages

Some conceptual/terminology variability that can be confusing in DDM

- ***Intuitively***, since the DDM is about two-choice decision, one would consider that one boundary represents choice A and the other choice B (this is sometimes called *stimulus coding*)



- In practice, *accuracy coding* is often used (when equivalent choices). It considers that one boundary is « correct response » and the other is « incorrect response »

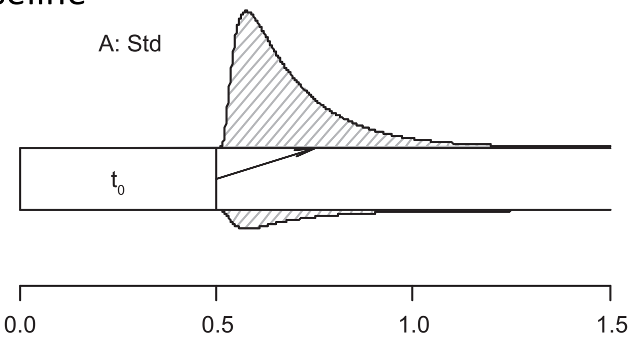


- It does not affect parameter estimation in most cases
- –> allows to aggregate more trials together to yield better estimates
- ***Except***, when you want evaluate if there is a bias in the starting point ($z \neq \frac{a}{2}$)
  - In these cases we must use *stimulus coding*

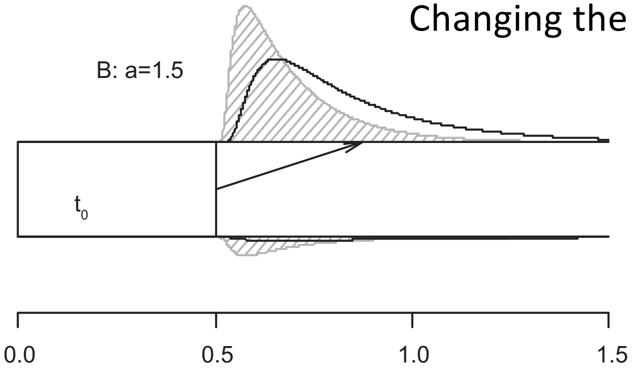
THE TOOLBOX WE'LL USE, USES THE INTUITIVE WAY ☺

Varying effects of parameters

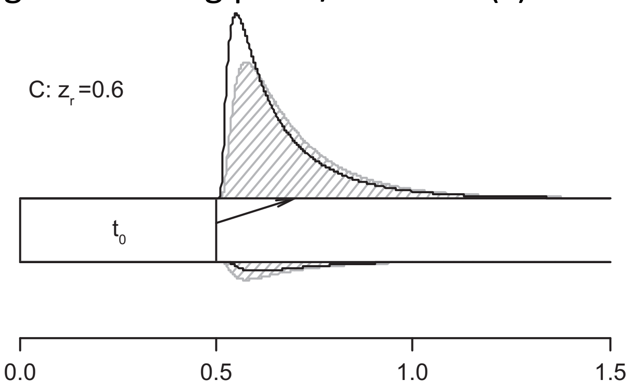
The baseline



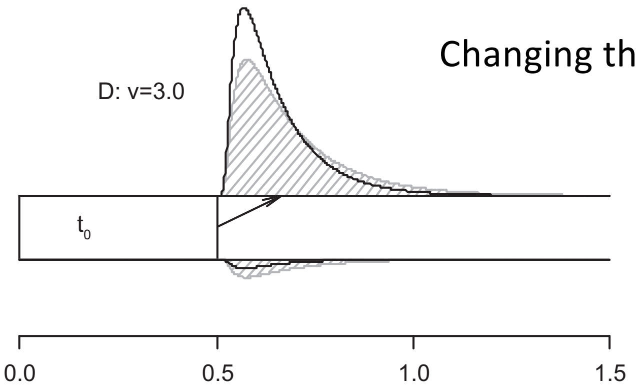
Changing the bound (a)



Changing the starting point/bias rate (z)



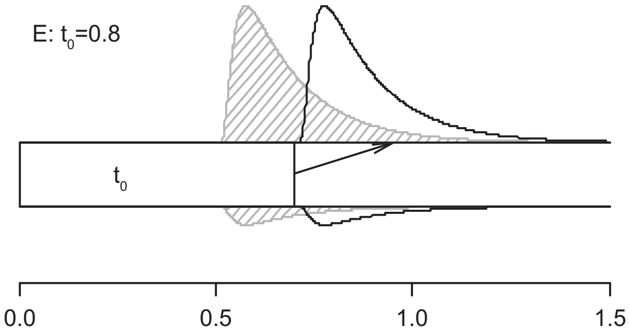
Changing the drift rate (v)



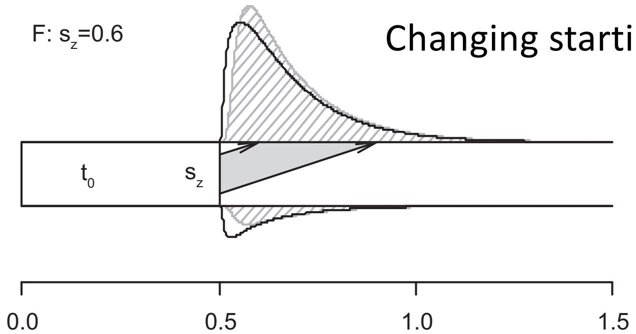
Voss et al., 2013

Varying effects of parameters

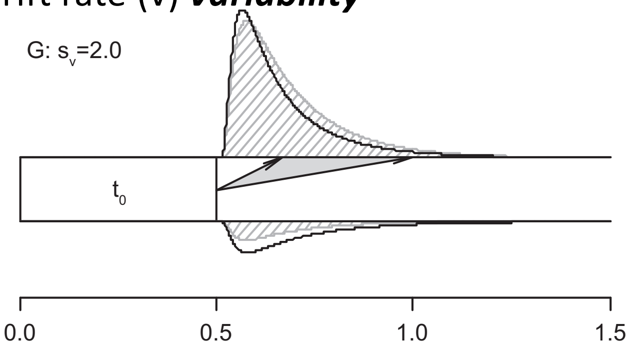
Changing Ter



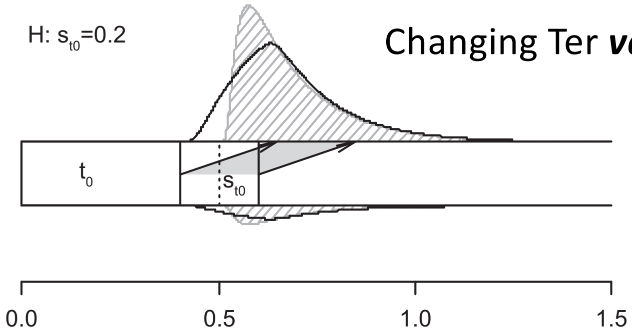
Changing starting point ***variability***



Changing drift rate (v) ***variability***



Changing Ter ***variability***

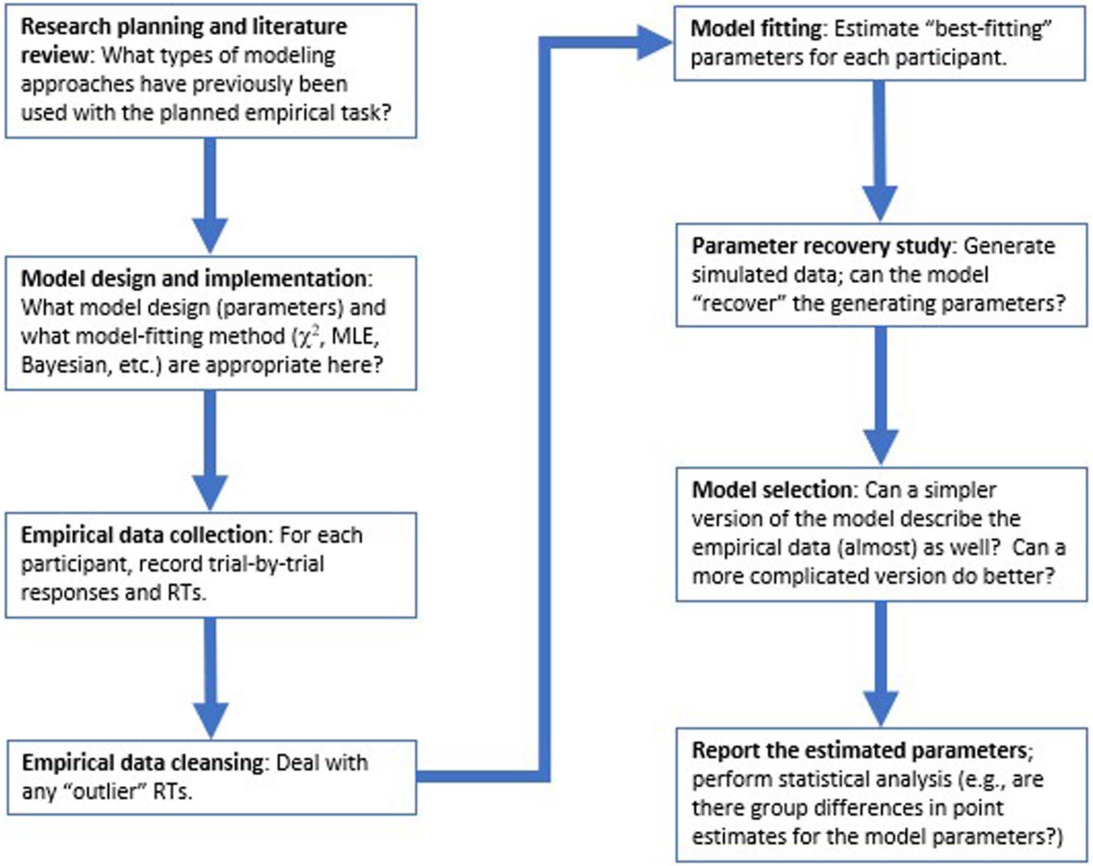


Voss et al., 2013

Fitting contest! by hand

Plan

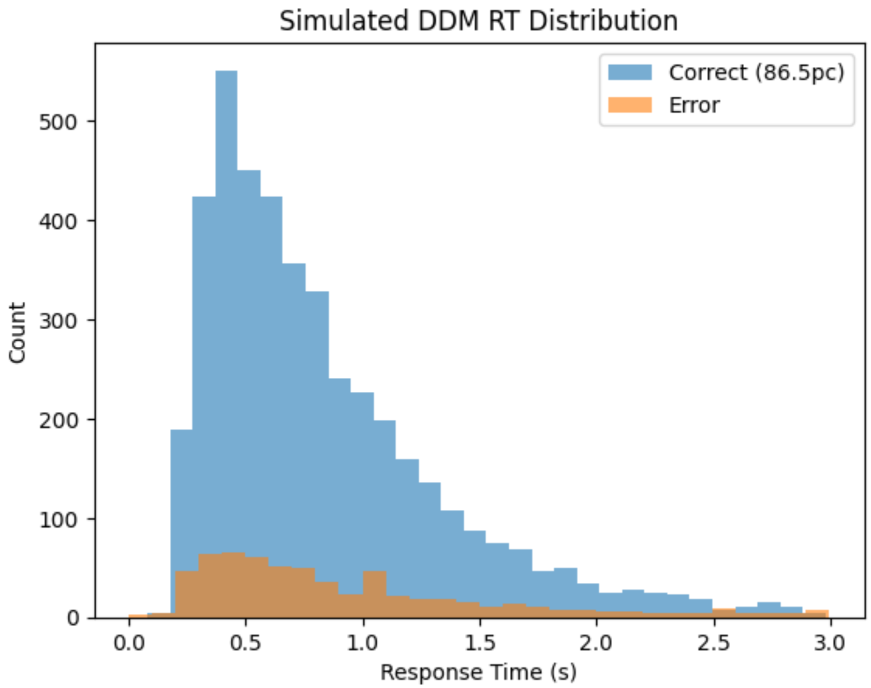
- The proper pipeline for a study using the DDM (using pyddm on colab)



Myers et al., 2022

STEP 1

- Set up a model with certain parameters
- Simulate a dataset
- Plot the simulations!



STEP 1

- Let’s look at the model summary

```
m.show()
```

```
Model information:
Choices: 'correct' (upper boundary), 'error' (lower boundary)
Drift component DriftConstant:
    constant
    Fixed parameters:
    - drift: 1.000000
Noise component NoiseConstant:
    constant
    Fixed parameters:
    - noise: 1.000000
Bound component BoundConstant:
    constant
    Fixed parameters:
    - B: 1.000000
IC component ICPointRatio:
    An arbitrary starting point expressed as a proportion of the distance between the bounds.
    Fixed parameters:
    - x0: 0.000000
Overlay component OverlayChain:
    Overlay component OverlayNonDecision:
        Add a non-decision by shifting the histogram
        Fixed parameters:
        - nondectime: 0.100000
    Overlay component OverlayUniformMixture:
        Uniform distribution mixture model
        Fixed parameters:
        - umixturecoef: 0.020000
```

STEP 2

- Let’s now try to recover the parameters we’ve used to simulate our data
- …

This is a (very small scale) parameter recovery: you’ve checked that when your model generated behavioral data with some parameters, it was itself able to « recognize » these parameters

STEP 3

- Load some real data
- Create a model in which only drift rate can vary (guess a bound)
  - Fit it and keep the BIC value
- Then a model in which only bound can vary (guess a drift rate)
  - Fit it and keep the BIC value
- Then a model in which both can vary
  - Fit it and keep the BIC value

This is a (very small scale) model comparison: you've checked which model best explained your data based on a corrected goodness of fit measure for each model

Example from a tutorial paper (that I recommend!)

	Estimated parameters, using DDM-4					*LLE* (from DDM-4)	*LLE* (from DDM-5)
	a	***z***	***d.s1***	***d.s2***	***Ter***		
1	1.08	0.36	−1.87	1.87	0.22	150.6	185.2
2	1.88	0.22	−2.35	2.35	0.24	49.9	70
3	1.74	0.22	−2.36	2.36	0.15	69.2	128.5
4	1.17	0.43	−2.51	2.51	0.17	273.6	287.1
5	1.08	0.36	−1.88	1.88	0.22	153	189.7
6	1.65	0.3	−2.42	2.42	0.25	140.5	140.5
7	1.2	0.36	−2.02	2.02	0.21	129.5	129.5
8	0.78	0.51	−1	1	0.23	220.4	220.4
9	0.83	0.59	−2.01	2.01	0.18	319.4	331.9
10	0.85	0.47	−0.45	0.45	0.18	128.9	135.6
Mean	1.23	0.38	−1.89	1.89	0.2	163.5	181.8
SD	0.4	0.12	0.67	0.67	0.03	84.7	79.6

(most current) DDM packages comparison

	PyDDM	HDDM	EZ-Diffusion	CHaRTr	DMAT	fast-dm
Language	Python3	Python2/3	Matlab, R, Javascript, or Excel	Requires both R and C	Matlab	Command line
Solver	Fokker-Planck, analytical	Analytical numerical hybrid	None	None (Monte Carlo)	Analytical numerical hybrid	Fokker-Planck
Task parameters						
Time dependence of drift/noise	Any function	Constant	Constant	Any function	Constant	Constant
Position dependence of drift/noise	Any function	Constant	Constant	Any function	Constant	Constant
Bounds	Any function	Constant	Constant	Any function	Constant	Constant
Parameter dependence on task conditions	Any relationship for any parameter	Regression model	Categorical	Categorical	Linear	Categorical

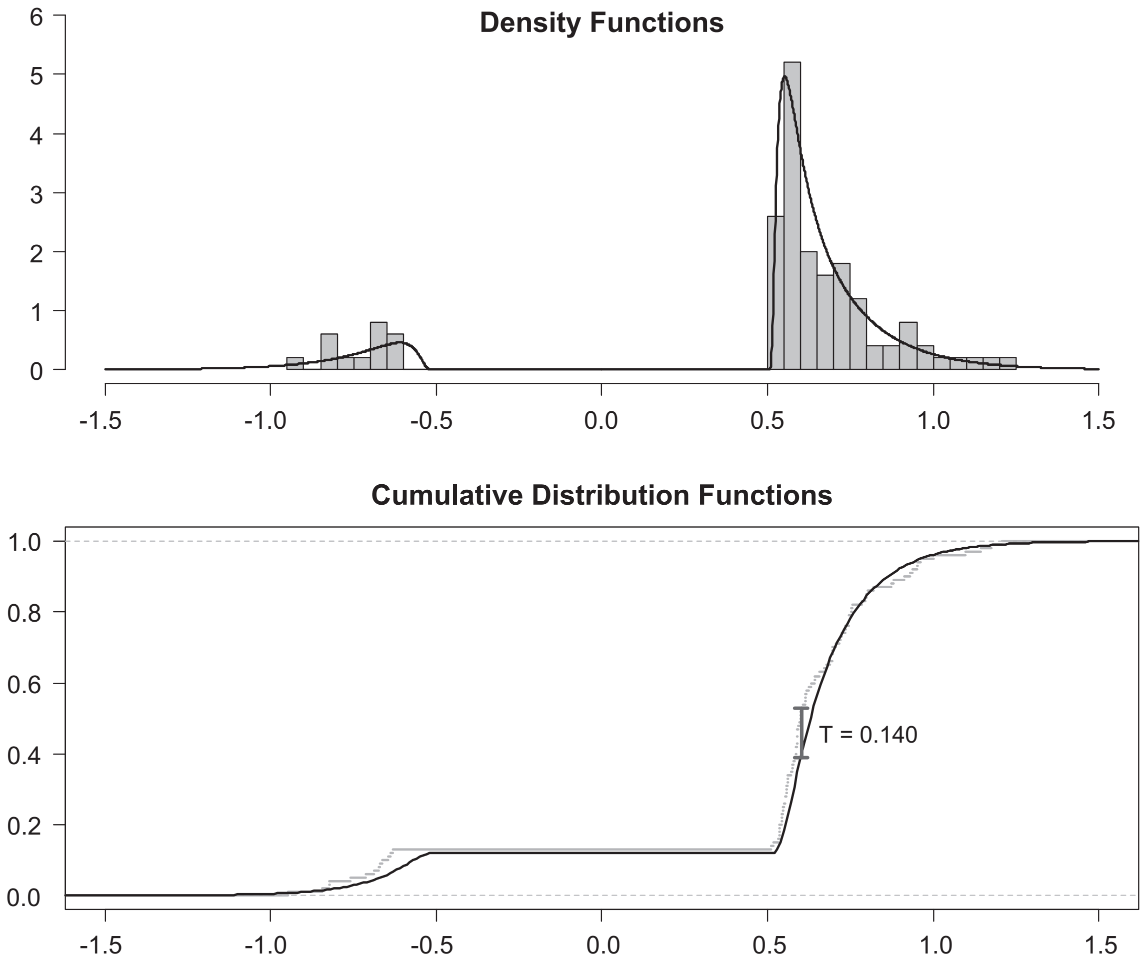
Shinn et al. (2020)

(most current) DDM packages comparison

	PyDDM	HDDM	EZ-Diffusion	CHaRTr	DMAT	fast-dm
Language	Python3	Python2/3	Matlab, R, Javascript, or Excel	Requires both R and C	Matlab	Command line
Across-trial variability						
Across-trial drift variability	Slow discretization (via extension)	Normal distribution	None	Any distribution	Normal distribution	Normal distribution
Across-trial starting point variability	Any distribution	Uniform distribution	None	Any distribution	Uniform distribution	Uniform distribution
Across-trial non-decision variability	Any distribution	Uniform distribution	None	Any distribution	Uniform distribution	Uniform distribution
Model simulation and fitting						
Hierarchical fitting	No	Yes	No	No	No	No
Fitting methods	Any numerical (default: differential evolution)	MCMC	Analytical	Any numerical	Nelder-Mead	Nelder-Mead
Objective function	Any function (default: likelihood)	Likelihood	Mean/stdev RT and P(correct)	Any sampled (e.g. quantile maximum likelihood)	Quantile maximum likelihood or chi-squared	Likelihood, chi-squared, Kolmogorov-Smirnov
Mixture model	Any distribution(s)	Uniform	None (extendable)	None	Uniform and undecided guesses	Uniform

Shinn et al. (2020)

How is the DDM optimized (very short version)

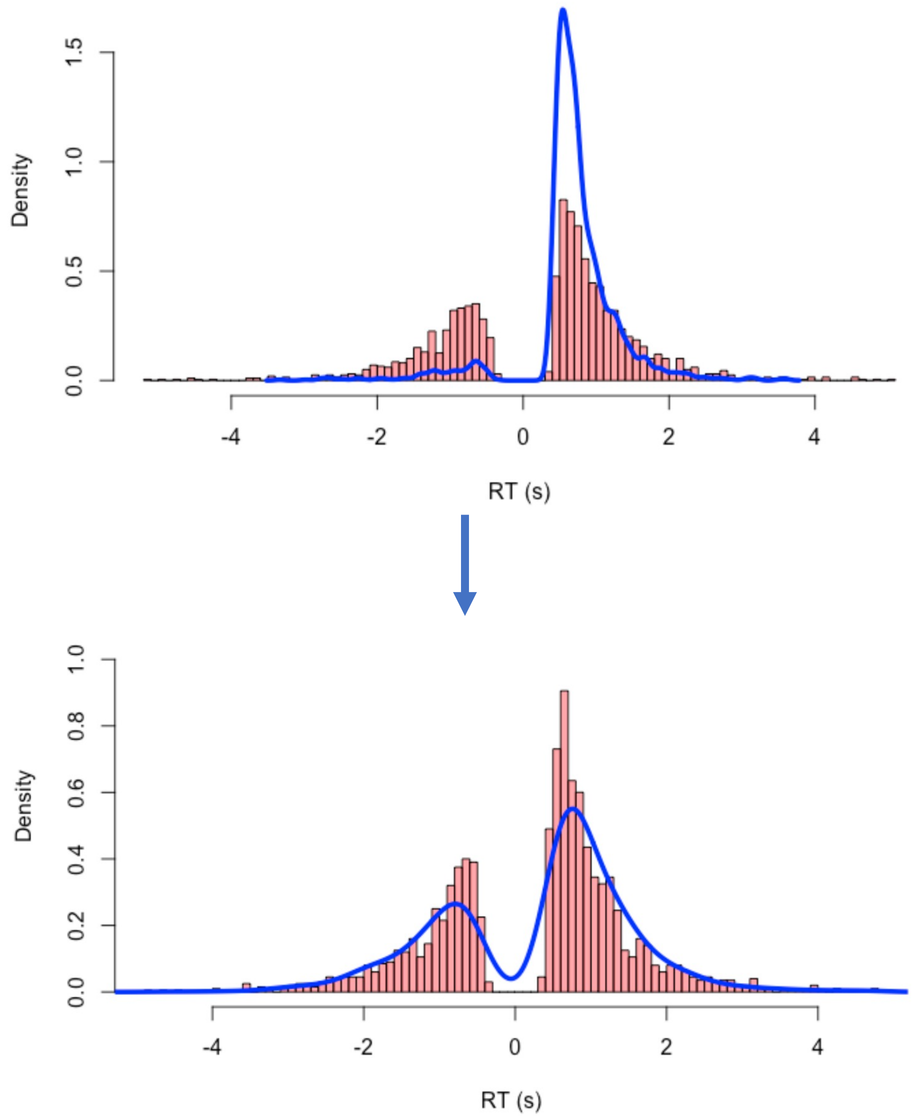


Voss et al, 2013

SLIDES FROM (the great) Luc Vermeylen

Objective Functions

- We need to find the parameters that best explain our data
    - An objective function quantifies the difference between our observed data and the data generated by some given candidate parameters
    - Next, an optimization algorithm can adjust the candidate parameters so that the misfit between observed and predicted data is minimized

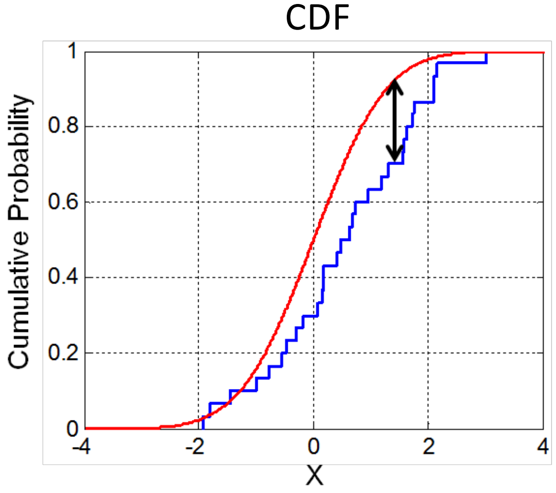


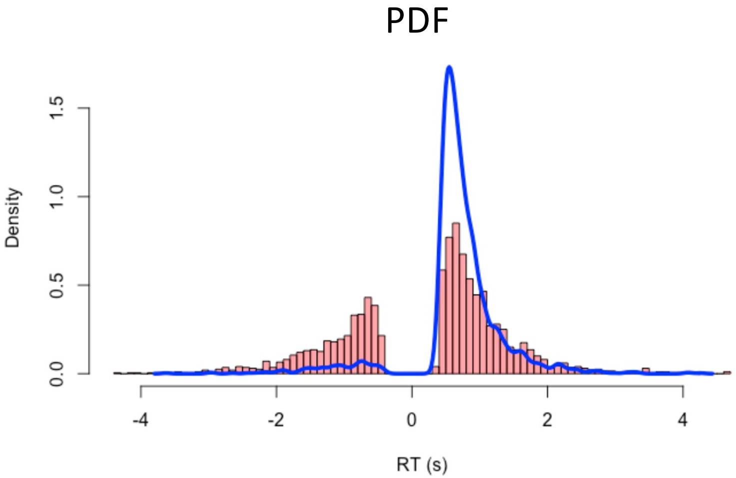
SLIDES FROM (the great) Luc Vermeylen

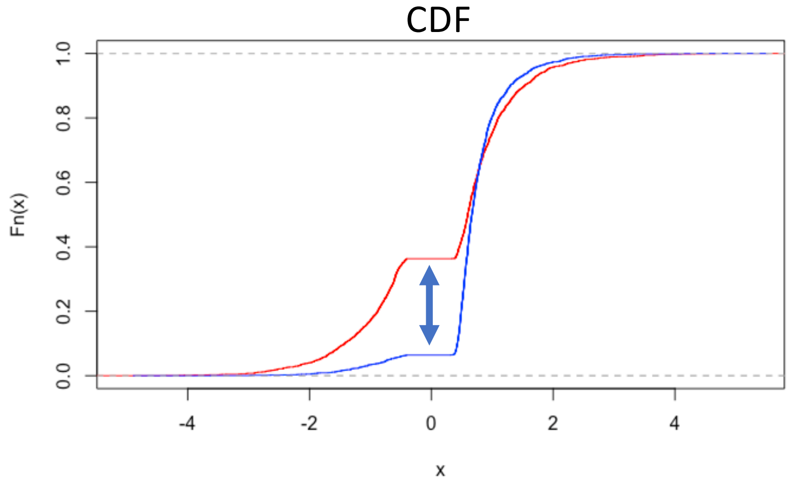
Kolmogorov-Smirnov (KS)

- The test statistic ("D") of the KS test is the maximum absolute difference between two cumulative distribution functions

Problem: we have two distributions… so we combine them!



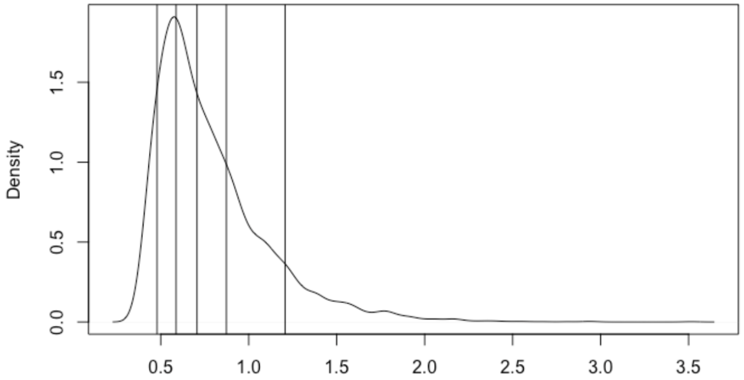




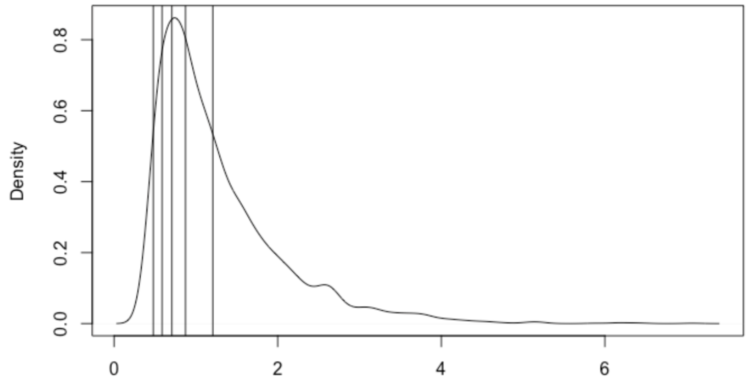
SLIDES FROM (the great) Luc Vermeylen

Multinomial Likelihood (G^2)

Compute proportion between quantiles (e.g., .1, .3, .5, .7, .9) on observed data
=> is by definition 10% 20% 20% 20% 20% 10%



What is the proportion between the quantiles (from the observed data) applied on the predicted data?



Next, quantify the mismatch:

G^2 = 2 * N * sum(observed_proportion * log(observed_proportion/predicted_proportion))

AIC = $-G^2$ + 2*k, where k, where k is the number of parameters
BIC = $-G^2$ + k*log(N), where N is the number of observations.

These are calles approximate AIC/BIC, because not based on the true likelihood.

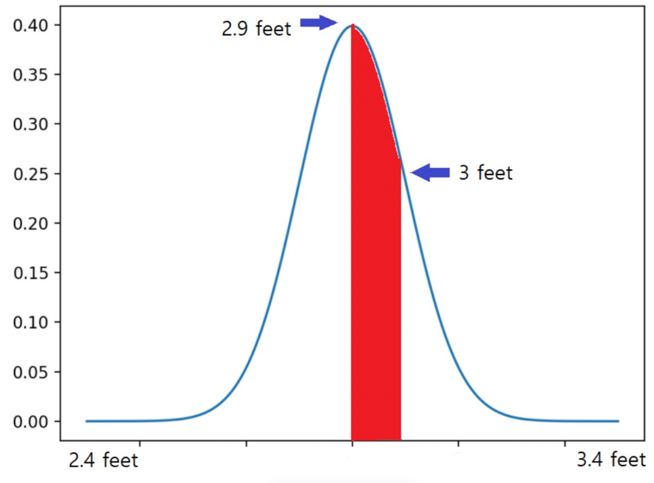
SLIDES FROM (the great) Luc Vermeylen

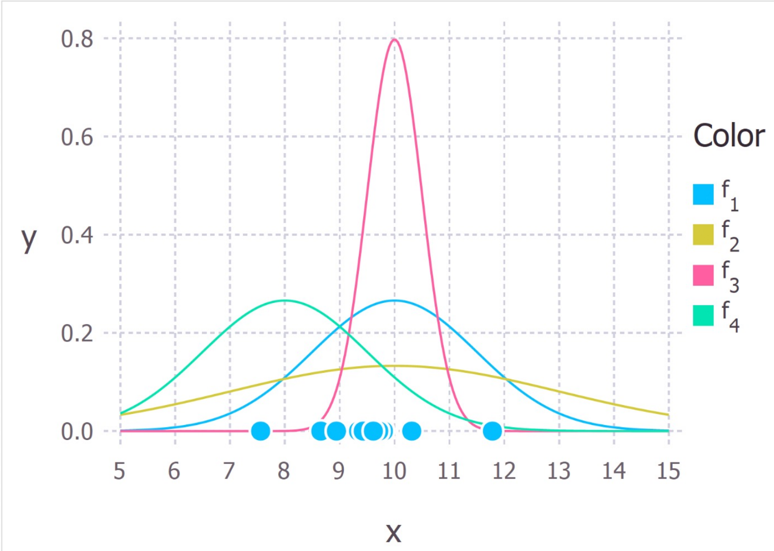
Maximum Likelihood (ML)

- **Probability: P(data|distribution)**
  - You ask a question about your data
  - **GIVEN** a distribution (e.g., N(0,1))
  - => the distribution is fixed, and ask questions about different sets of data
  - = Area Under the Curve (AUC) and sums to 1

- **Likelihood: L(distribution|data)**
  - You ask a question about your distribution
  - **GIVEN** a set of data
  - => the data is fixed, and you ask questions about different sets of distributions
  - Does not sum to 1 (higher is simply better fit)

- **Maximum Likelihood:**
  - You shift the parameters of the distribution
  - So that you maximize the likelihood of the data
  - "find the distribution that best explains the data"

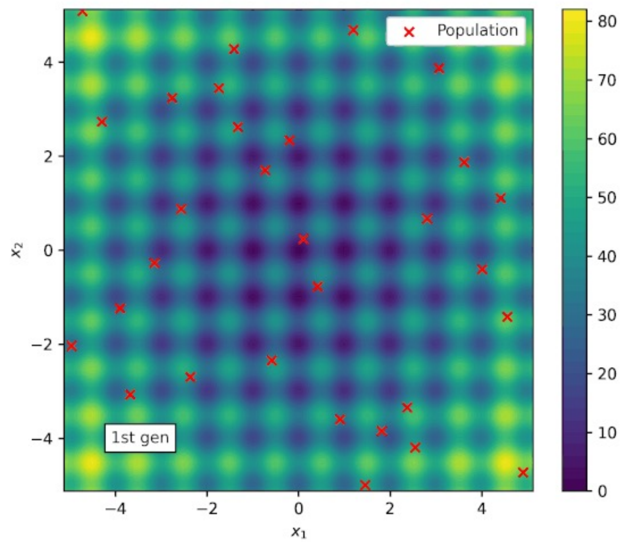




SLIDES FROM (the great) Luc Vermeylen

Differential Evolution (DE) Optimization

- A very powerful optimization method that can solve high-dimensional problems
- the DE algorithm works by having a population of candidate solutions (called agents)
- moved around in the search-space by using simple mathematical formulae to combine the positions of existing agents from the population
- If the new position of an agent is an improvement then it is accepted and forms part of the population, otherwise the new position is simply discarded
- The process is repeated and by doing so it is hoped, but not guaranteed, that a satisfactory solution will eventually be discovered.



front-end models

- task-specific function acts as an initial filter for the stimulus information before it feeds into the *back-end* EAM (Evidence Accumulation Model) process
  - e.g. a front-end model could extract some specific visual features from the stimuli, combine information from different modalities, etc.

2020 ■ Vol. 16 ■ no. 2

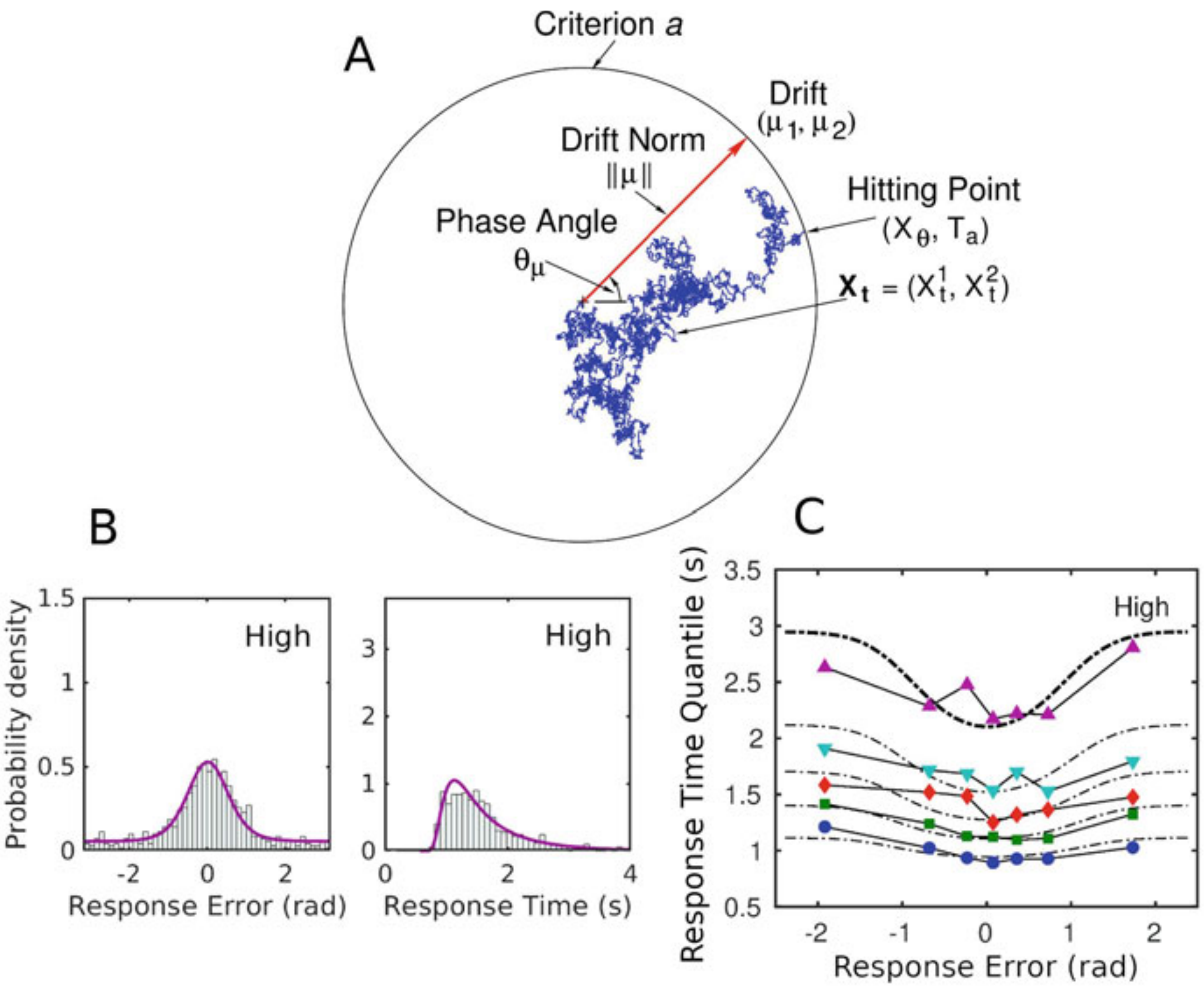
When accuracy rates and mean response times lead to false conclusions: A simulation study based on the diffusion model

Veronika Lerche [a] and Andreas Voss [a]

[a] Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany

« [...] our results show that qualitatively different sets of diffusion model parameters can lead to the same pattern of these aggregated variables, which renders the interpretation of these *[behavioral (RT, accuracy)]* measures ambiguous. »

« Often, these inferences will hold; however, this will not always be the case because—as we demonstrated in this article based on a simulation study—qualitatively different sets of model parameters can result in the same mean RTs and accuracy rates. »


A
Criterion a
Drift
(μ_1, μ_2)
Drift Norm
$\|\mu\|$
Hitting Point
(X_θ, T_a)
Phase Angle
θ_μ
$\mathbf{X_t} = (X_t^1, X_t^2)$
B
Probability density
High
Response Error (rad)
High
Response Time (s)
C
Response Time Quantile (s)
High
Response Error (rad)

Ratcliff &

DDM *works*

- Participants instructed to “work especially carefully and avoid mistakes,” show larger boundary separation a (corresponding to greater response caution (Voss et al., 2004)
- Participants working under time pressure show reduced boundary separation a (Milosavljevic et al., 2010)
- When one response is more frequent or more highly rewarded, the starting point z shifts to favor that response (Ratcliff and McKoon, 2008; Mulder et al., 2012; Arnold et al., 2015)
- When stimulus discriminability is varied, making the task harder or easier, the drift rate d changes accordingly (Ratcliff and McKoon, 2008)
    - Similar effect when stimulus difficulty is constant but participants are deprived of sleep for 24 h (Johnson et al., 2021).
- Introducing a motor response difficulty (e.g. requiring multiple keypresses for each response (Lerche and Voss, 2019), increases T_{er}

DDM *uncovers invisible effects in classic behavioral measures*

- **Distinguishing Progressive supranuclear palsy and Parkinson's disease cognitive deficits in a Go/No-Go task (Zhang et al., 2016)**

  « Mechanistic differences underlying participants' poor performance were not observable from classical analysis of behavioural data, but were clearly revealed by modelling. »

  « These differences provide a rational basis on which to develop and assess new therapeutic strategies for cognition and behaviour in these disorders. »

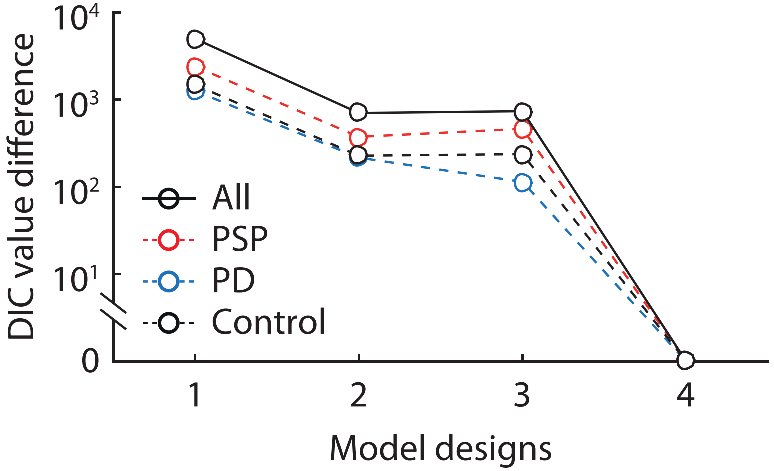
doi:10.1093/brain/awv331 BRAIN 2016: 139; 161–173 | 161

BRAIN
A JOURNAL OF NEUROLOGY

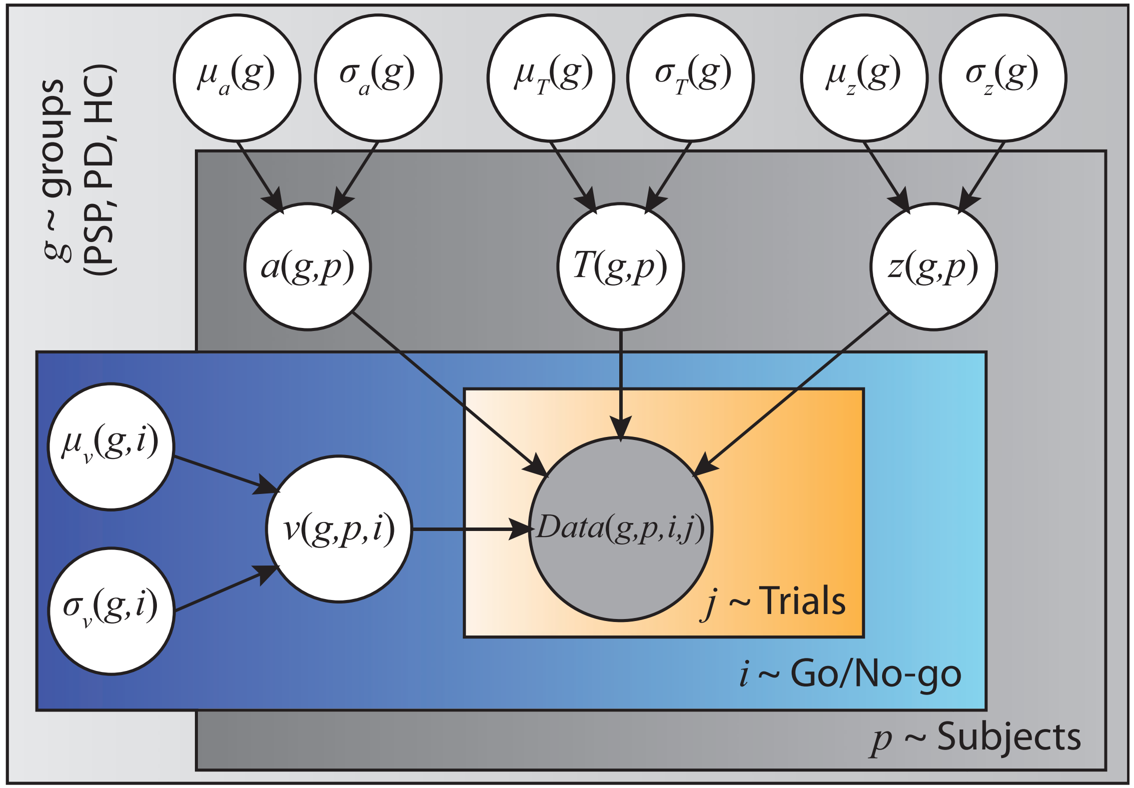
Different decision deficits impair response inhibition in progressive supranuclear palsy and Parkinson's disease

Jiaxiang Zhang,[1,2] Timothy Rittman,[3] Cristina Nombela,[3] Alessandro Fois,[3]
Ian Coyle-Gilchrist,[3] Roger A. Barker,[3] Laura E. Hughes[2,3] and James B. Rowe[2,3,4]

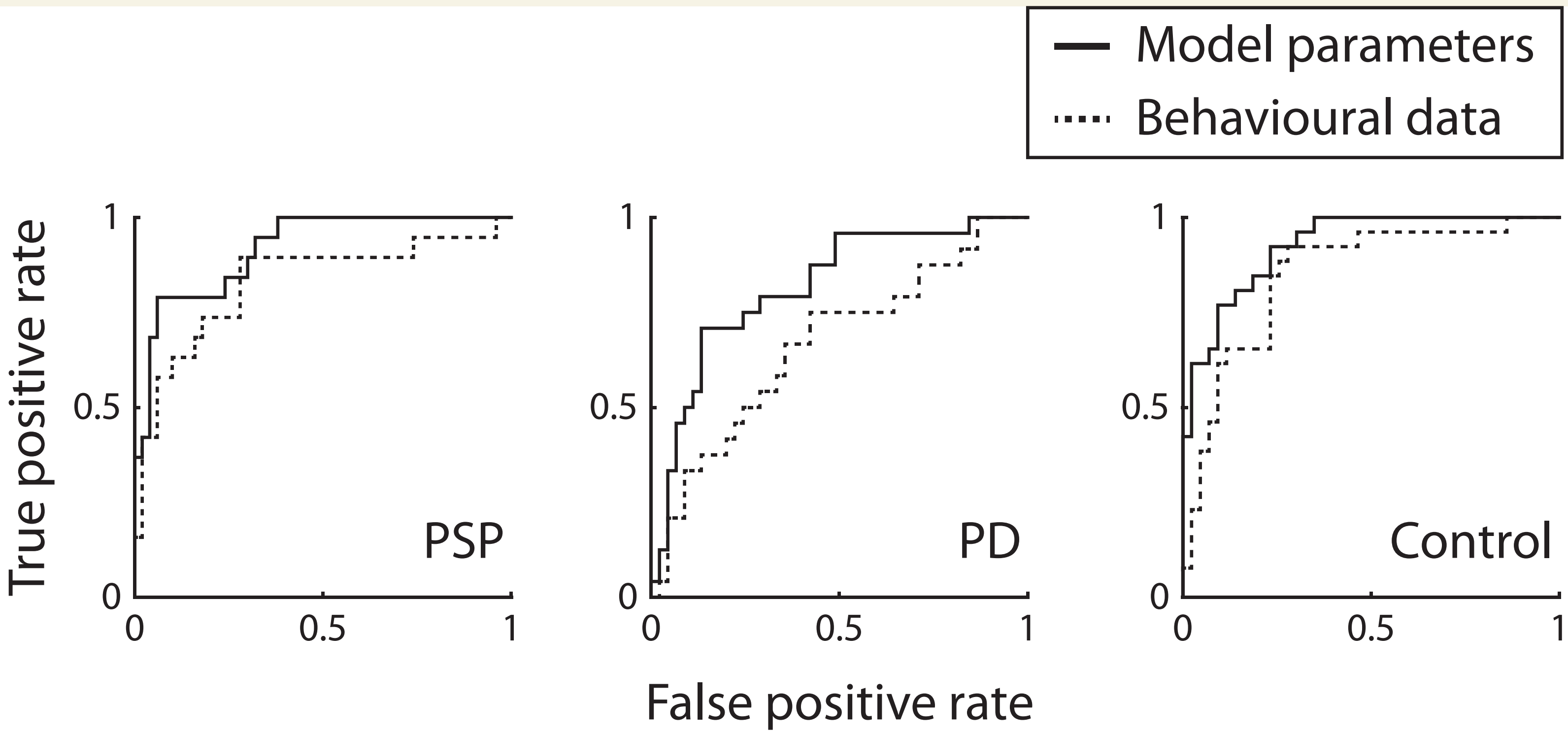
DDM *uncovers invisible effects in classic behavioral measures*



B

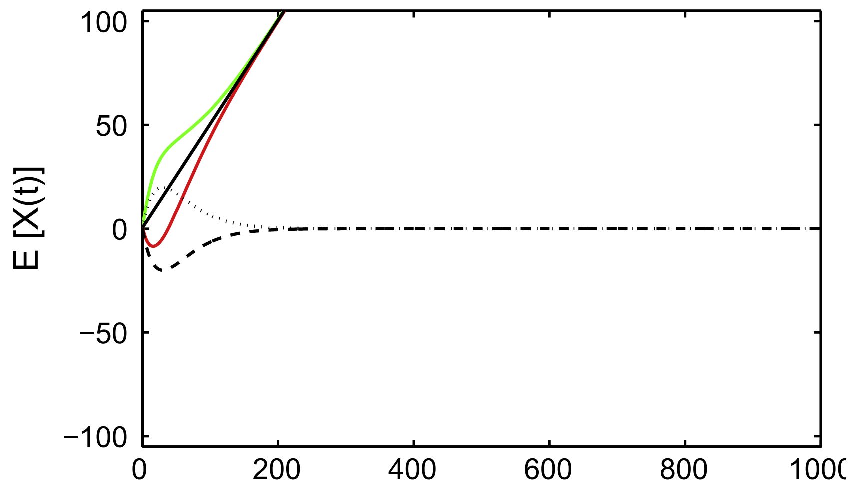


DDM *uncovers invisible effects in classic behavioral measures*

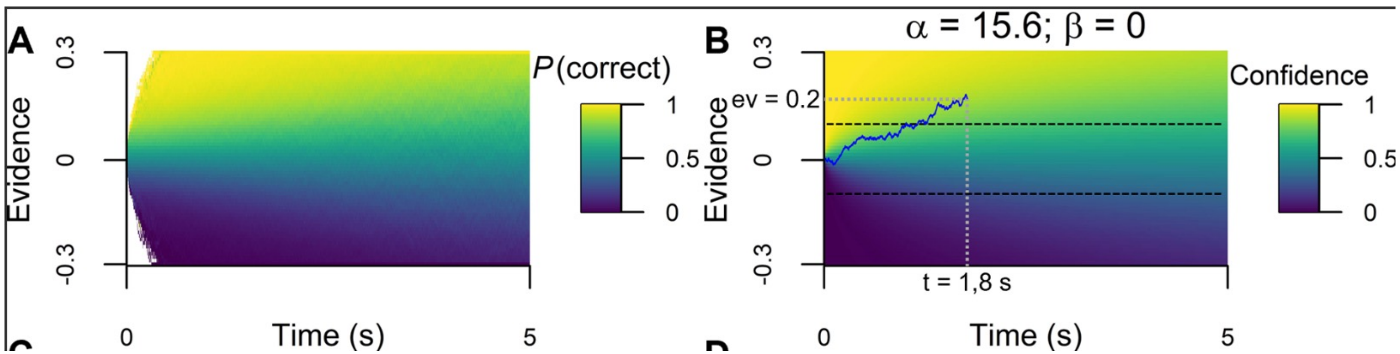


« Add-ons »

- Conflict DDM
  (Ulrich et al, 2015; Lee & Sewell, 2024)



- Confidence & DDM (Le Denmat et al, 2024)



PyDDM Cookbook

Below are recipies for how to get specific models and features to work in PyDDM. If you have not yet worked through the Quickstart guide, please do that first! This will give you the fundamentals for understanding the models below, demonstrated using leaky integration, collapsing bounds, and a coherence-dependent drift rate.

Common models

- Attention DDM (ADDM)
- Reinforcement learning DDM (RL-DDM)
- DDM with parameter variability ("full DDM")
- Weibull collapsing bounds
- Multi-sensory integration

Model components

- Biased drift rate
- Changing drift rate
- Non-linear drift rate
- Unique moment-to-moment drift rate on each trial (e.g., matching EEG)
- Biased starting position
- Starting position variability
- Non-decision time variability
- Urgency signal as a multiplicative gain
- Attractor states

Take home messages

- The DDM is more or less the gold standard in (binary) decision making