Quantifying The Benefits Of Modeling Decision Making Over Measuring Speed And Accuracy Alone.

Tom Stafford, Angelo Pirrone
Parallelisation: Mike Croucher
Shiny app: Anna Krystalli

These slides: goo.gl/FbuJEb

Psychonomics Amsterdam, Berlagezaal,
Thursday 10th May 2018, 14:00 - 14:20

The Speed-Accuracy Trade Off (SATO)

Fast but inaccurate vs Slow but accurate

- Individual differences: definitely
- Group differences: likely
- RT and accuracy contain different information on the decision process (Palmer, Huk & Shadlen, 2005; Stone, 2014)

This choice for participants becomes a confound for psychologists

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. Journal of Vision, 5 (5), 1–1.

Stone, J. V. (2014). Using reaction times and binary responses to estimate psychophysical performance: An information theoretic analysis. Frontiers in Neuroscience, 8, 35.

The Speed-Accuracy Trade Off (SATO)

Popular (but bad) solutions

1. Ignore either speed or accuracy

    e.g. psychophysics

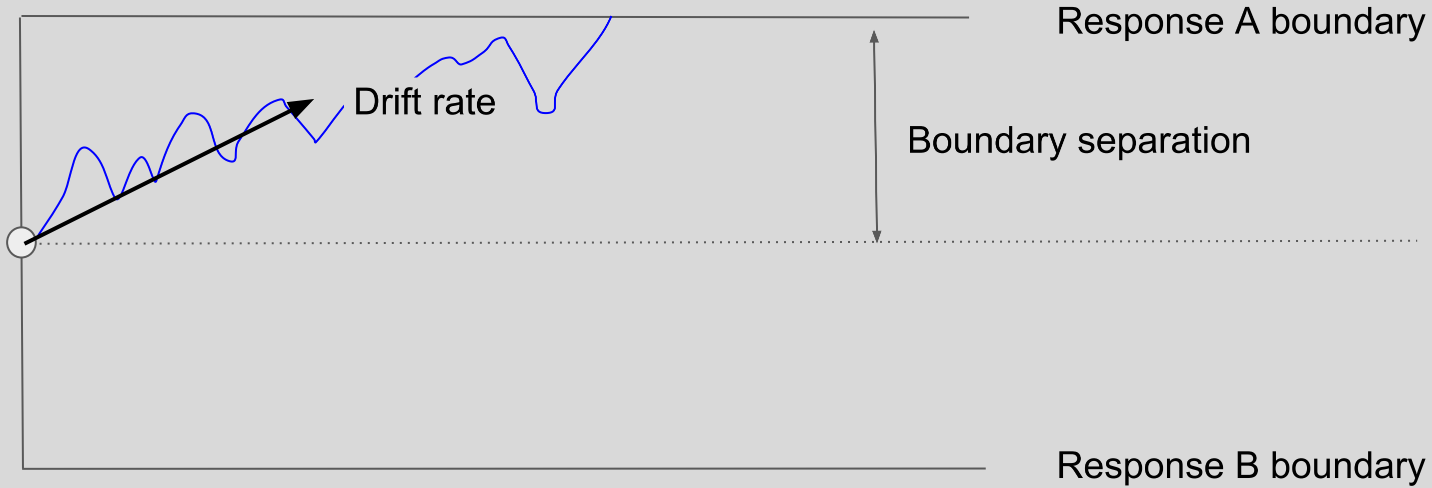
2. Ignore one after failing to find a significant difference on the other

3. Linearly combine

    e.g. ANOVA models, inefficiency scores (RT/accuracy)

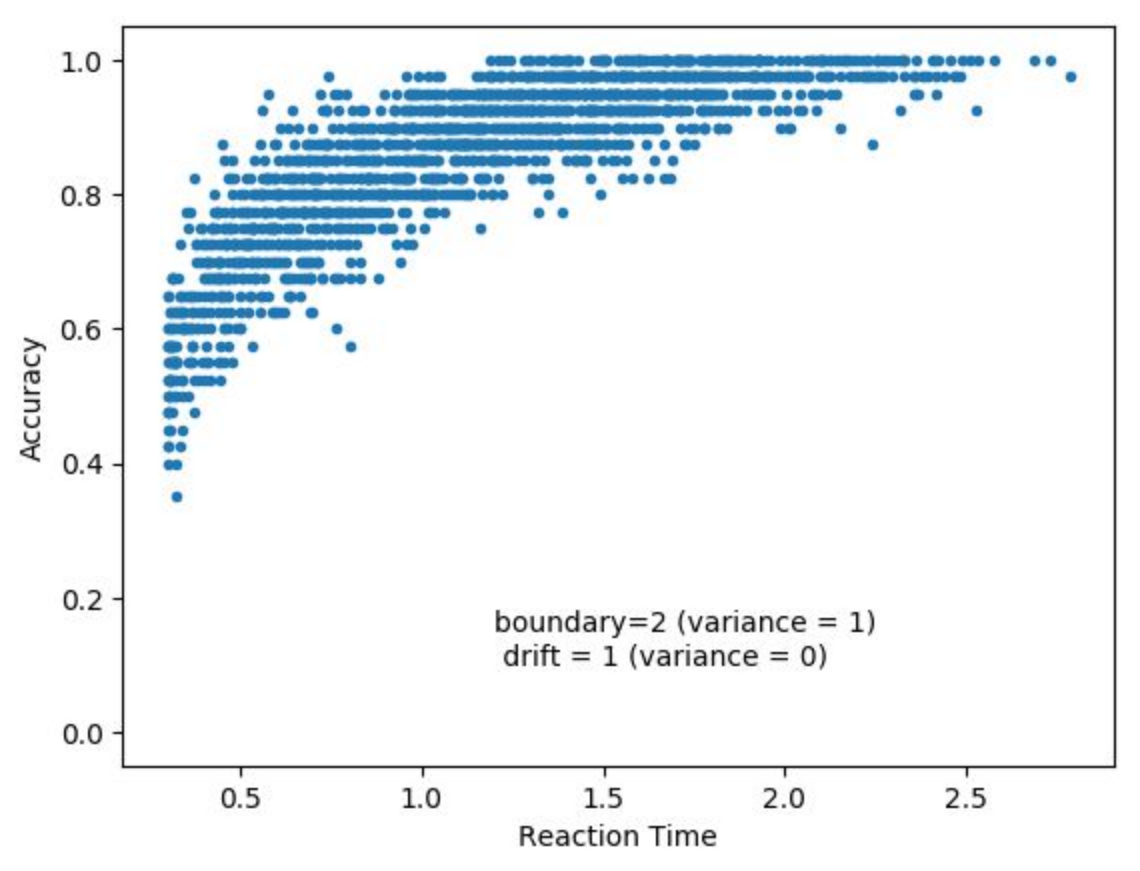
Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (ies) a better dependent variable than the mean reaction time (rt) and the percentage of errors (pe)? Psychologica Belgica, 51 (1), 5–13.
Townsend, J.T., & Ashby, F.G. (1983). Stochastic modeling of elementary psychological processes. Cambridge: Cambridge University Press

Decision models



Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review, 85*(2), 59-108
Smith, P. L., and Ratcliff, R.(2004) Psychology and neurobiology of simple decisions. *Trends in neurosciences 27*(3), 161-168.

Visualising the Speed-Accuracy trade-off



1.0
0.8
0.6
0.4
0.2
0.0
Accuracy
0.5
1.0
1.5
2.0
2.5
Reaction Time
boundary=2 (variance = 1)
drift = 1 (variance = 0)

Simulation strategy...simulate & test group differences

Define for 2 groups

Drift
Boundary

→

Generate

RTs
Accuracy

→

Model

(Estimates of)
Drift

Test group differences

Simulation strategy...repeat many times

Simulated expt and parameter recovery

Sensitivity measures for drift, RT and Accuracy

Group difference detected or not?

x 2000

Hit Rate, False Alarms, d’

Technical Details

Code available here : github.com/tomstafford/ddm_sims

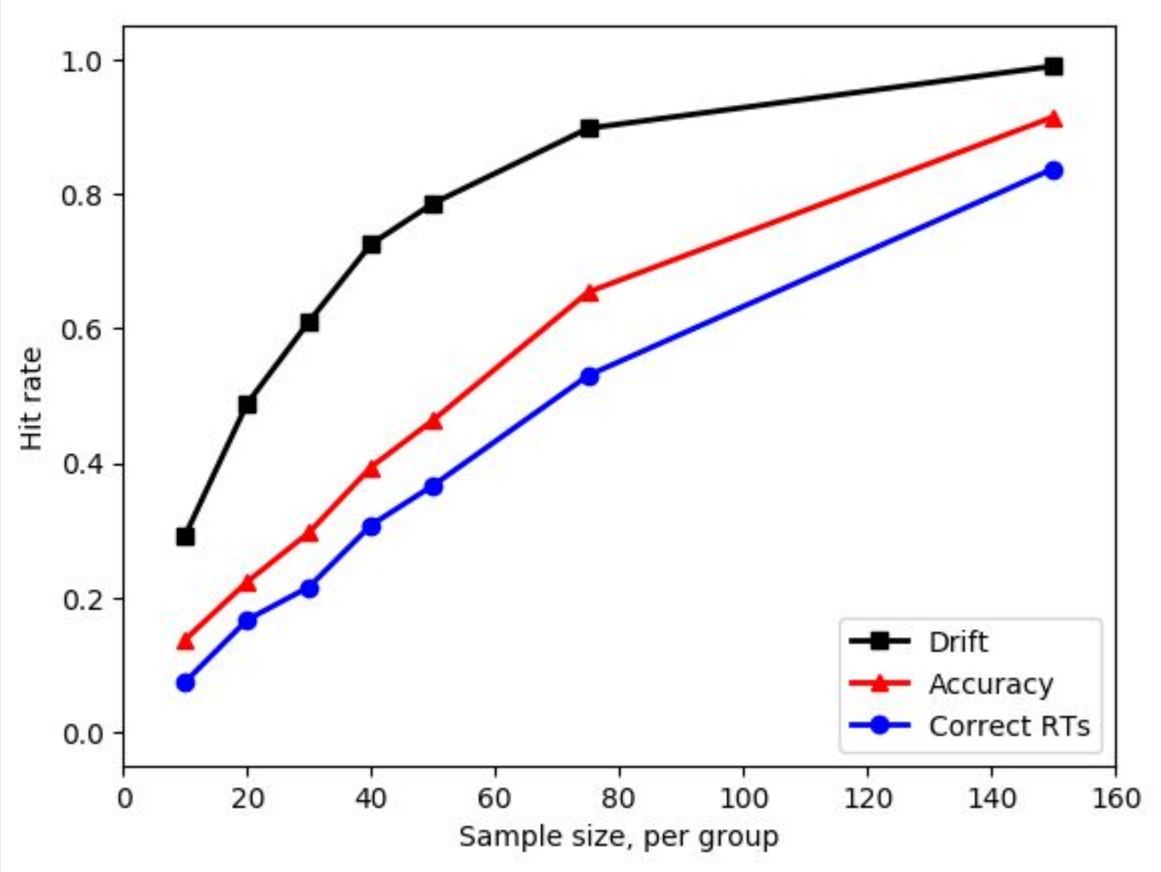
Decision modelling using HDDM (Wiecki, Sofer & Frank, 2013)

Parallelisation implemented by Mike Croucher, Co-founder Research Software Engineering at University of Sheffield (now Director of Research Computing, University of Leeds) - thanks Mike!

Results qualitatively similar if done using fast-dm (Voss & Voss, 2007)

Voss, A., & Voss, J. (2007). Fast-dm: A Free Program for Efficient Diffusion Model Analysis. *Behavioral Research Methods, 39,* 767-775
http://www.psychologie.uni-heidelberg.de/ae/meth/fast-dm/index.htm
Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical bayesian estimation of the drift-diffusion model in python. Frontiers in Neuroinformatics, 7, 14.

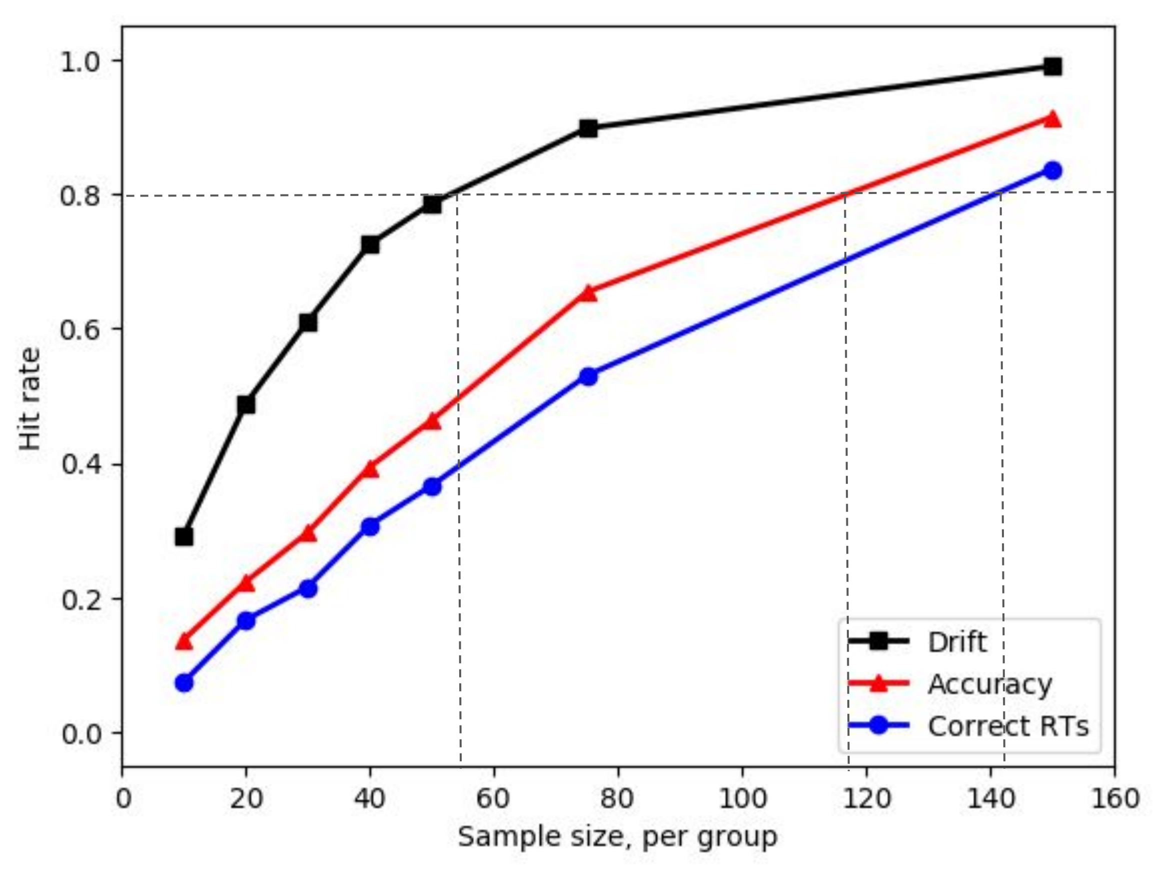
Results



Hit Rate

No Sato

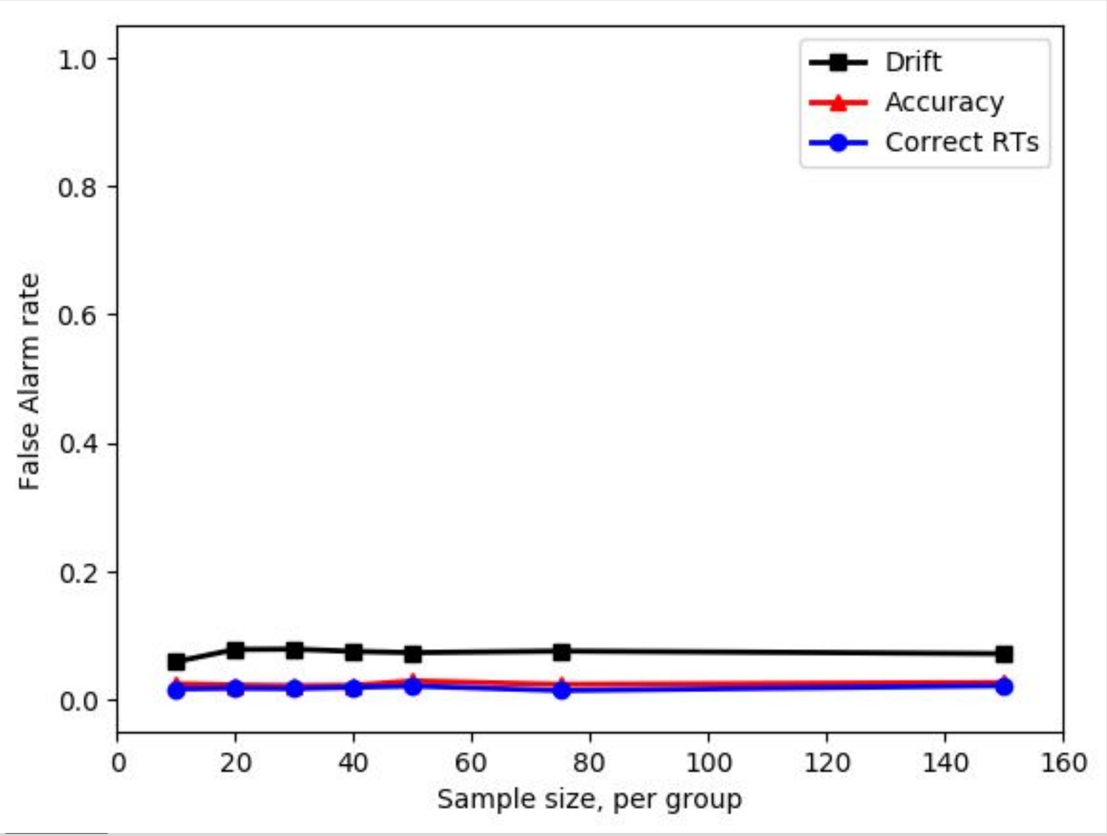
Parameters:
drift: 1 vs 1.1
boundary: 2 vs 2
intersubj var = 0.05
trials/ppt = 40



ppts/group for 80% power:

RT: ~140
Acc: ~115
Drift: ~55

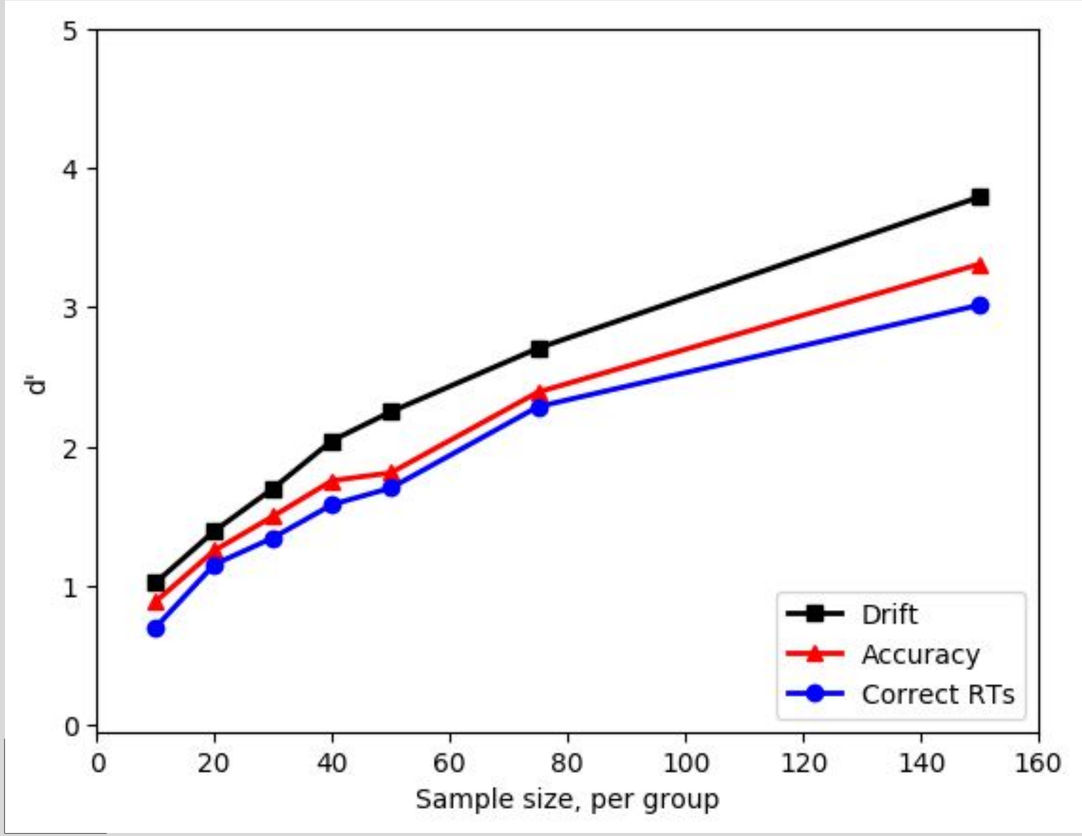
trials/ppt = 40



False Alarms

No Sato

Parameters:
drift: 1 vs 1.1
boundary: 2 vs 2
intersubj var = 0.05
trials/ppt = 40

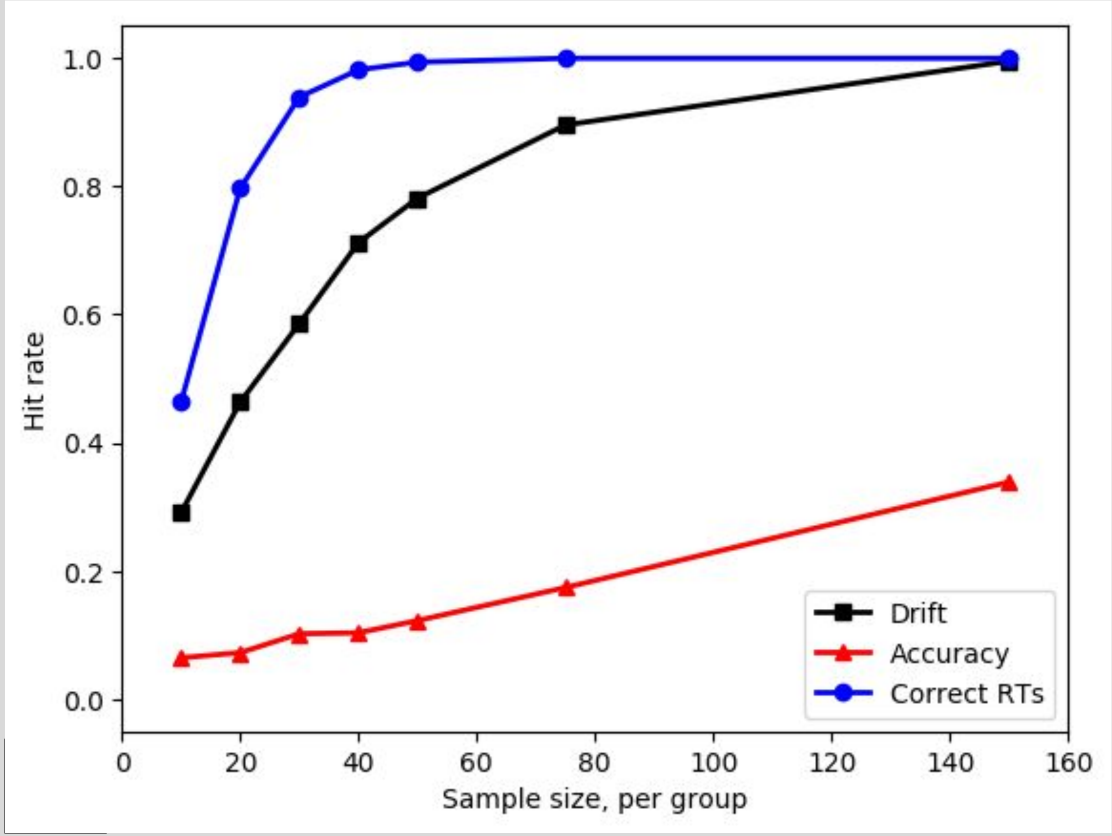


d’

No Sato

Parameters:
drift: 1 vs 1.1
boundary: 2 vs 2
intersubj var = 0.05
trials/ppt = 40

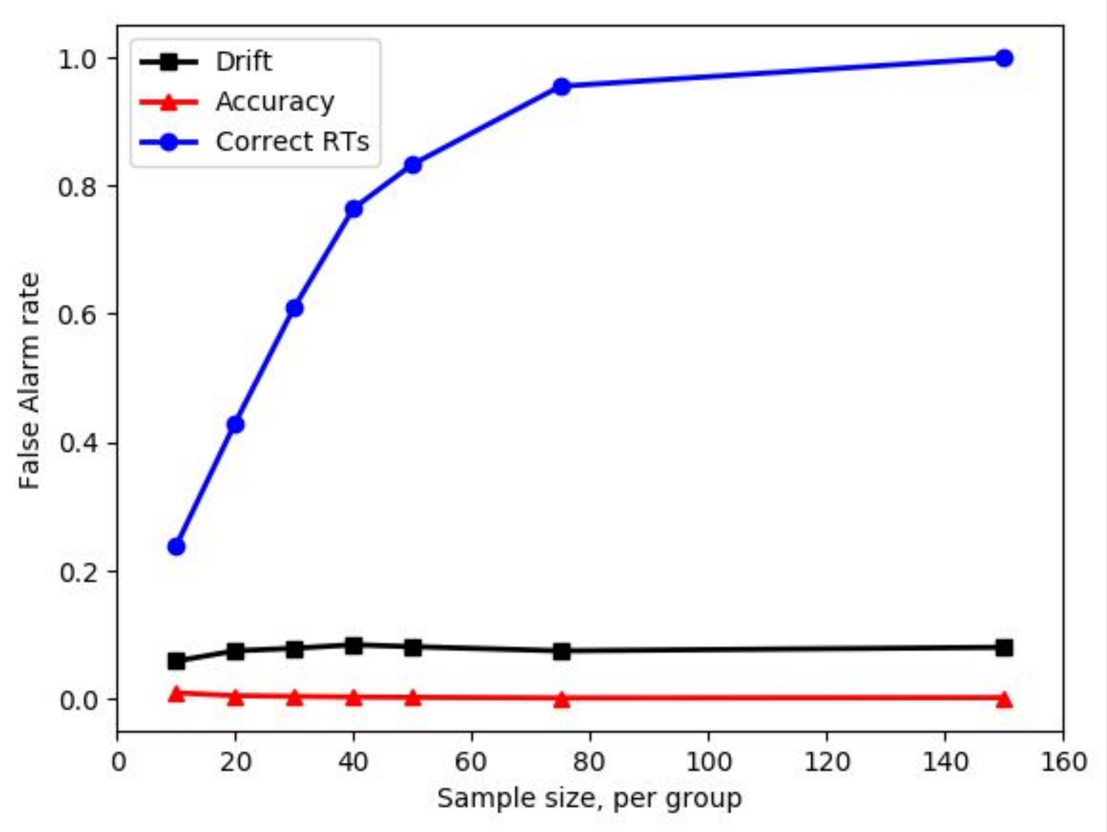
With Speed-Accuracy Trade Off
(i.e. between group boundary shift)



Hit Rate

With Sato

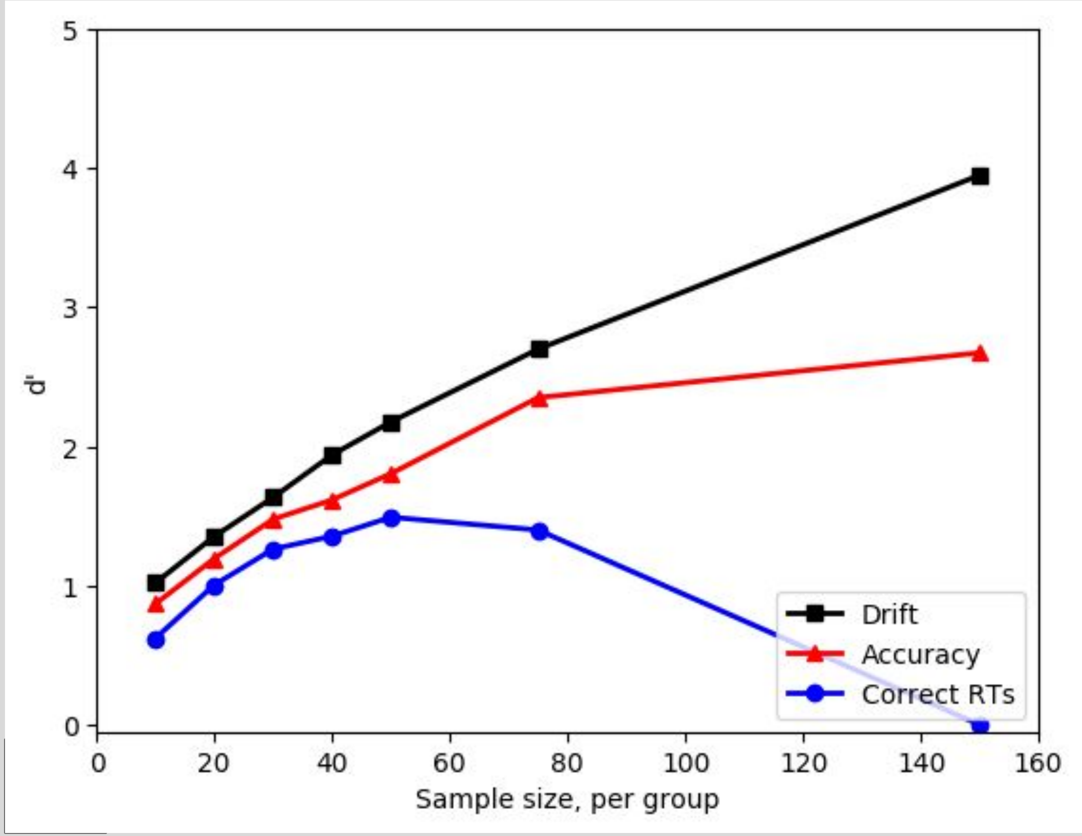
Parameters:
drift: 1 vs 1.1
boundary: 2 vs 1.9
intersubj var = 0.05
trials/ppt = 40



False Alarms

With Sato

Parameters:
drift: 1 vs 1.1
boundary: 2 vs 1.9
intersubj var = 0.05
trials/ppt = 40

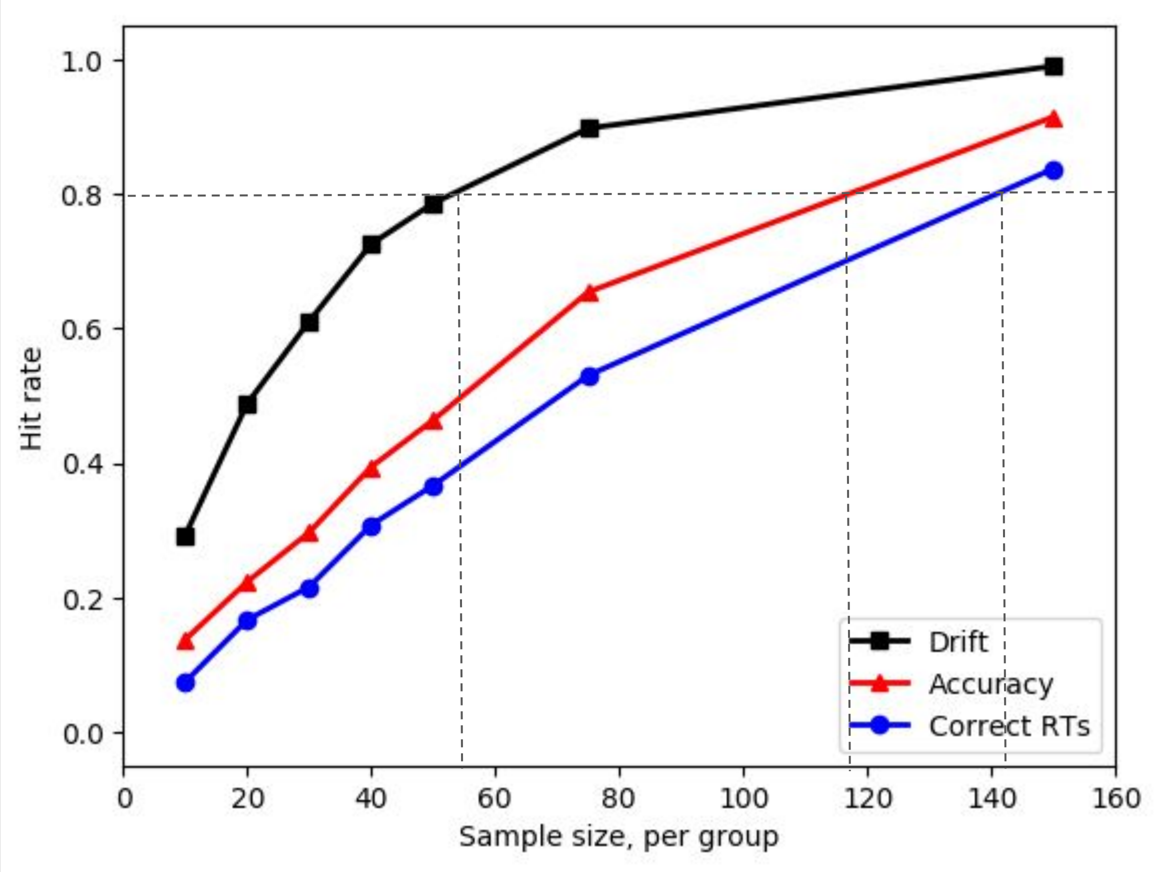


d’

With Sato

Parameters:
drift: 1 vs 1.1
boundary: 2 vs 1.9
intersubj var = 0.05
trials/ppt = 40

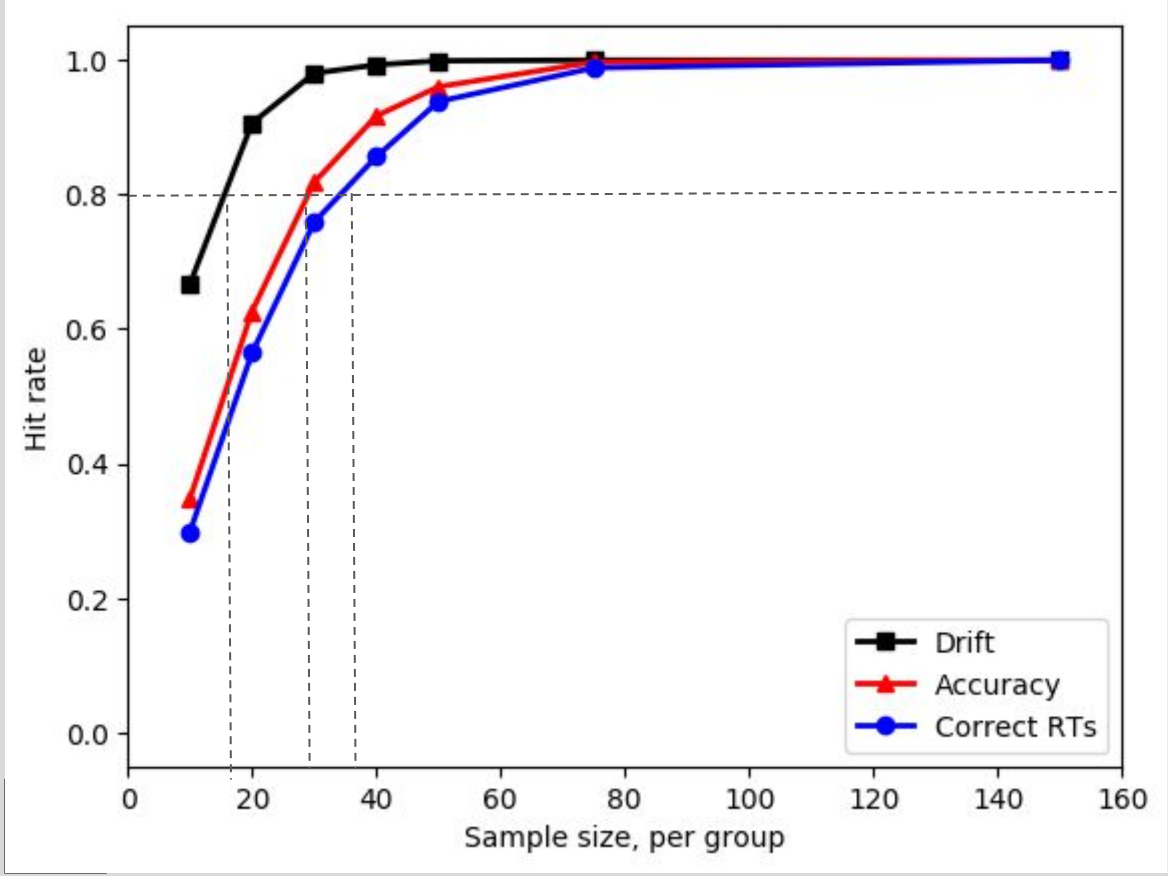
No SATO, but larger true effect



ppts/group for 80% power:

RT: ~140
Acc: ~115
Drift: ~55

trials/ppt = 40

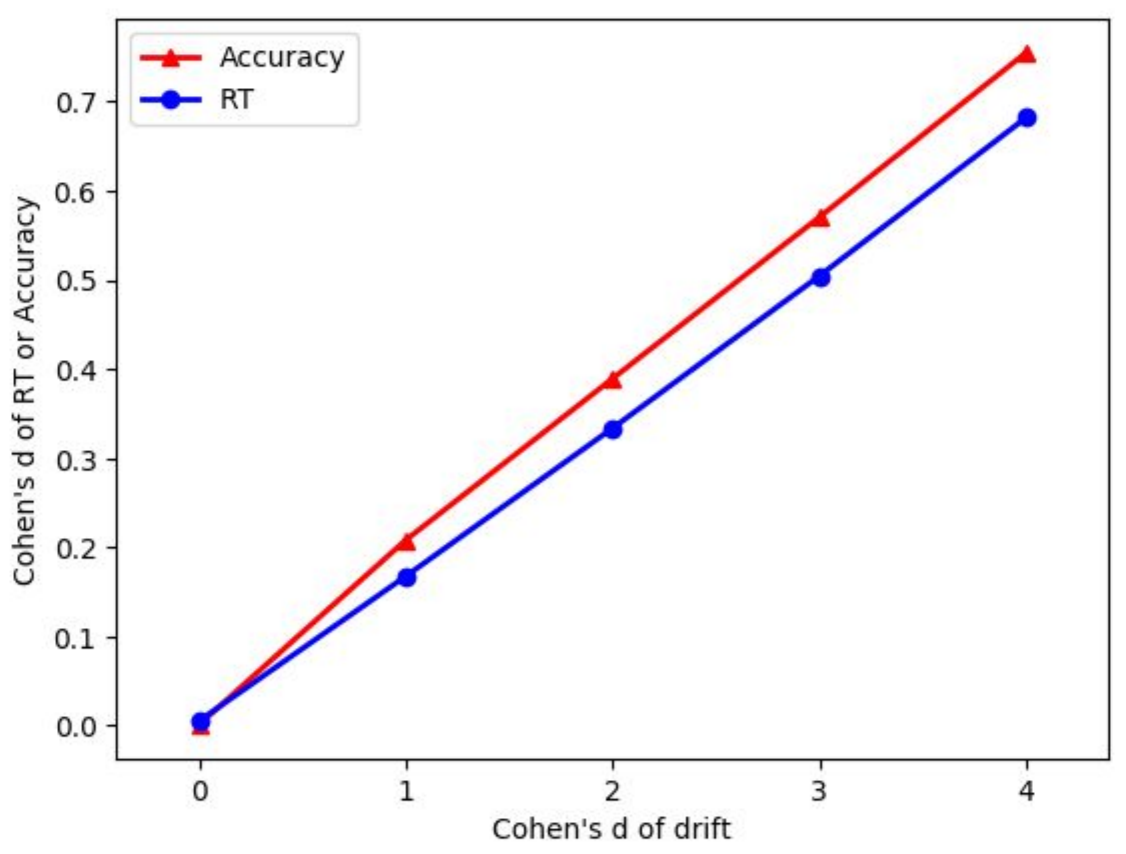


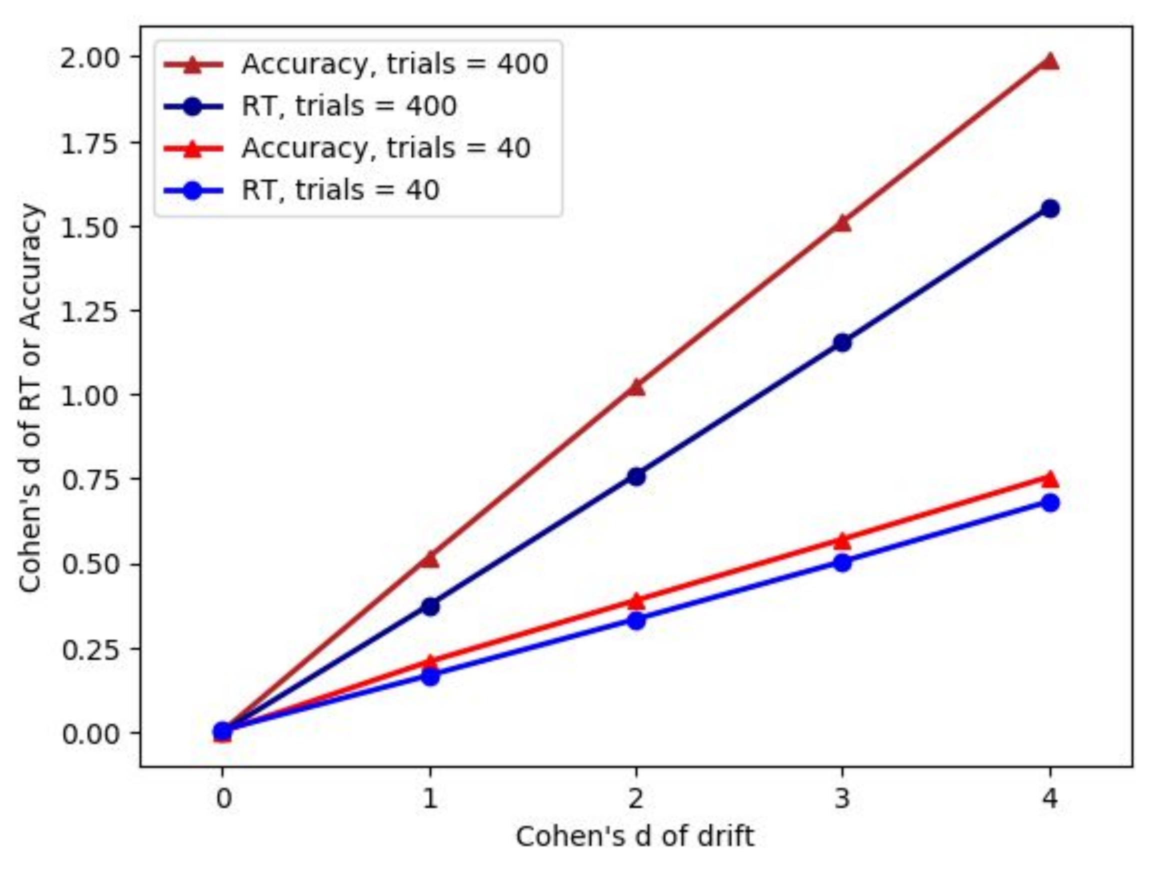
ppts/group for 80% power, larger effect:

RT: ~38
Acc: ~30
Drift: ~18

trials/ppt = 400

Why accuracy better than RT?


Accuracy
RT
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
2
3
4
Cohen's d of RT or Accuracy
Cohen's d of drift


Accuracy, trials = 400
RT, trials = 400
Accuracy, trials = 40
RT, trials = 40
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Cohen's d of RT or Accuracy
0
1
2
3
4
Cohen's d of drift

Massive benefits of decision modelling

- Power gain over analysing reaction time or accuracy alone
  - commonly allows <50% subjects for same power
- Avoid false positives due to speed-accuracy trade-offs
  - example: Pirrone, A., Dickinson, A., Gomez, R., Stafford, T. and Milne, E. (2017). [Understanding perceptual judgement in autism spectrum disorder using the drift diffusion model](). *Neuropsychology, 31* (2), 173-180

Future work

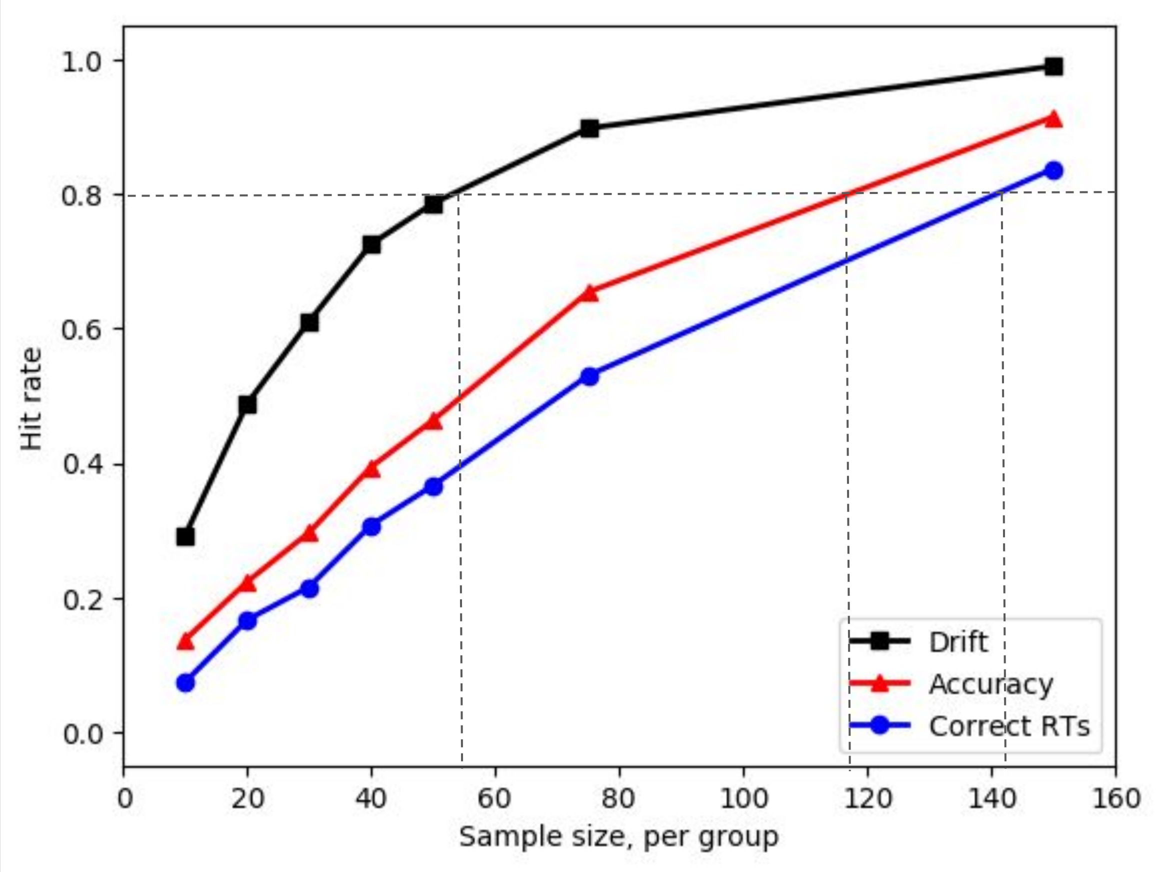
- Generalise work to different decision models and parameter regimes
- Interactive tool for exploring experiment design

https://github.com/tomstafford/ddm_sims

t.stafford@shef.ac.uk

END

Slides beyond this for reference only

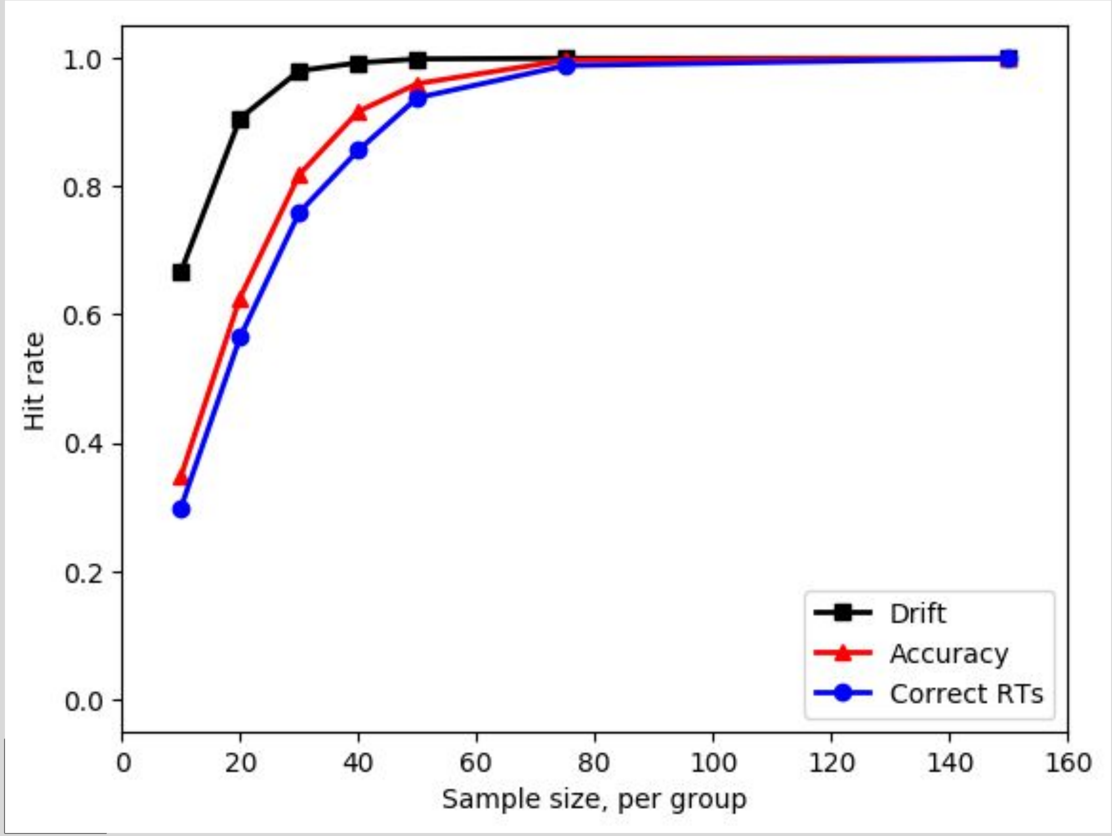


ppts/group for 80% power:

RT: ~140
Acc: ~115
Drift: ~55

trials/ppt = 40

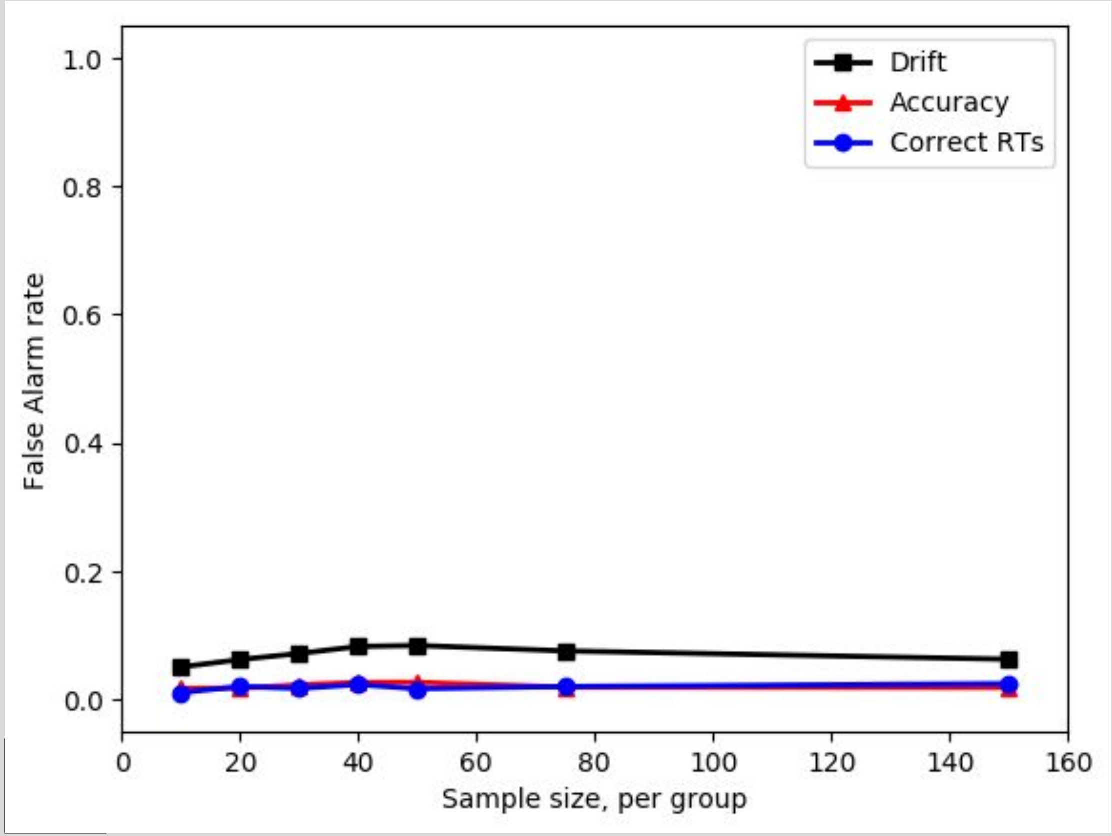
No SATO, Large drift effect (= 4)



Hit Rate

No Sato, larger effect

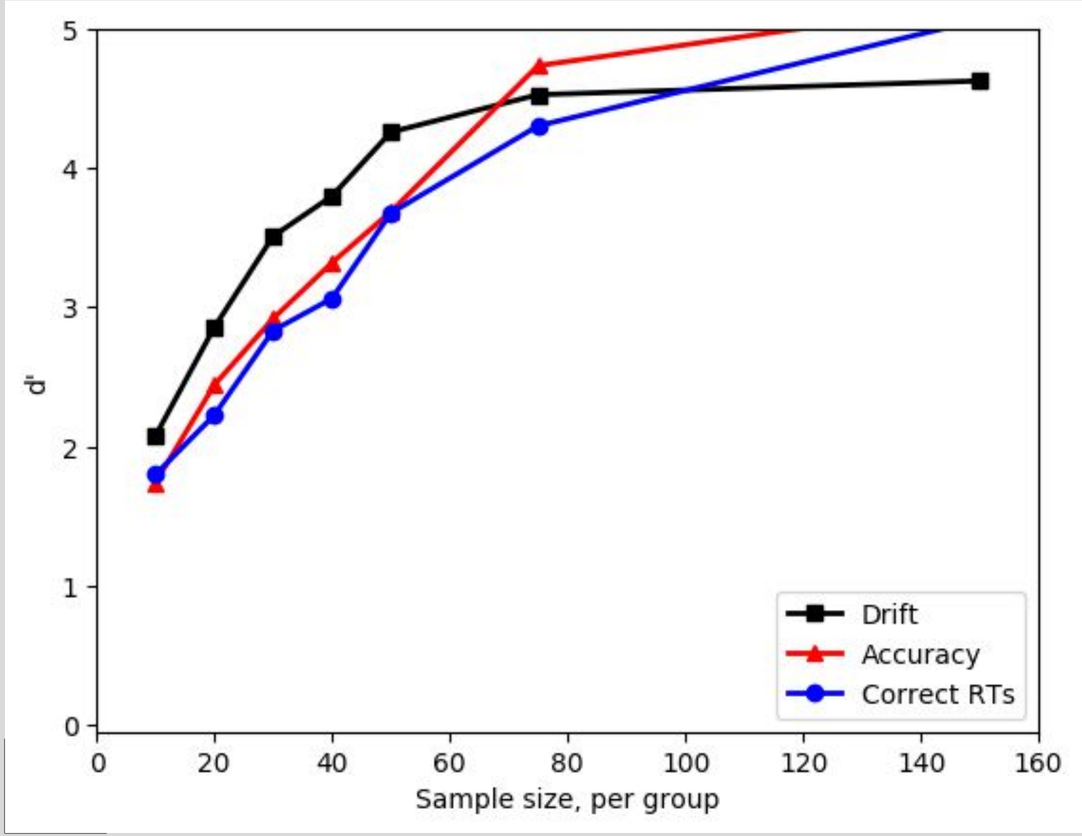
Parameters:
drift: 1 vs 1.2
boundary: 2 vs 2
intersubj var = 0.05
trials/ppt = 40



False Alarms

No Sato, larger effect

Parameters:
drift: 1 vs 1.2
boundary: 2 vs 2
intersubj var = 0.05
trials/ppt = 40

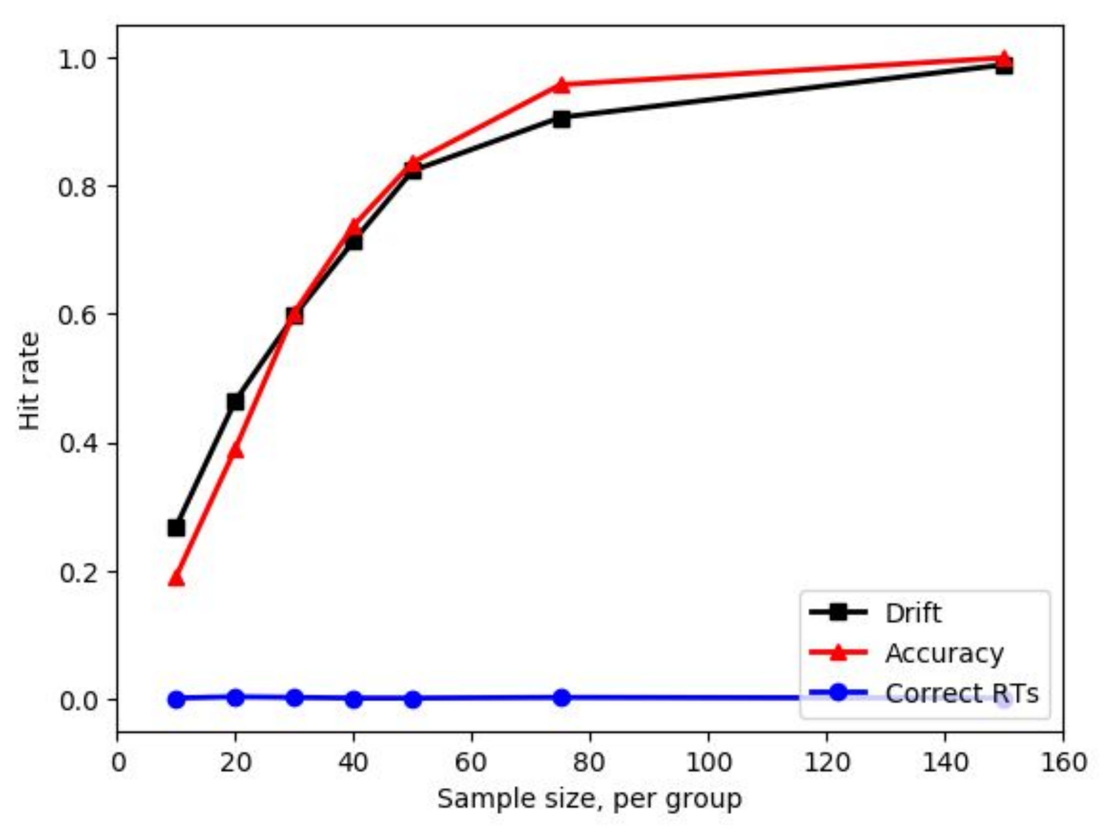


d’

No Sato, larger effect

Parameters:
drift: 1 vs 1.2
boundary: 2 vs 2
intersubj var = 0.05
trials/ppt = 40

SATO, Boundary shift up



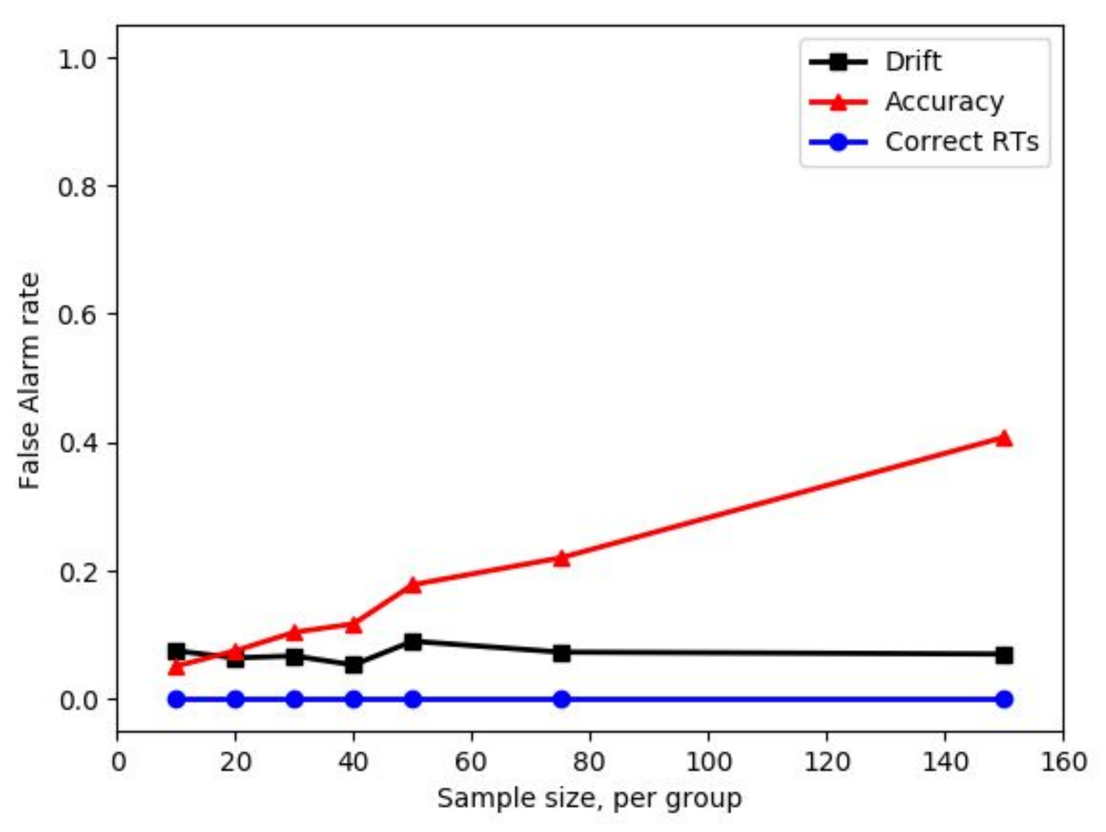
Hit Rate

With Sato

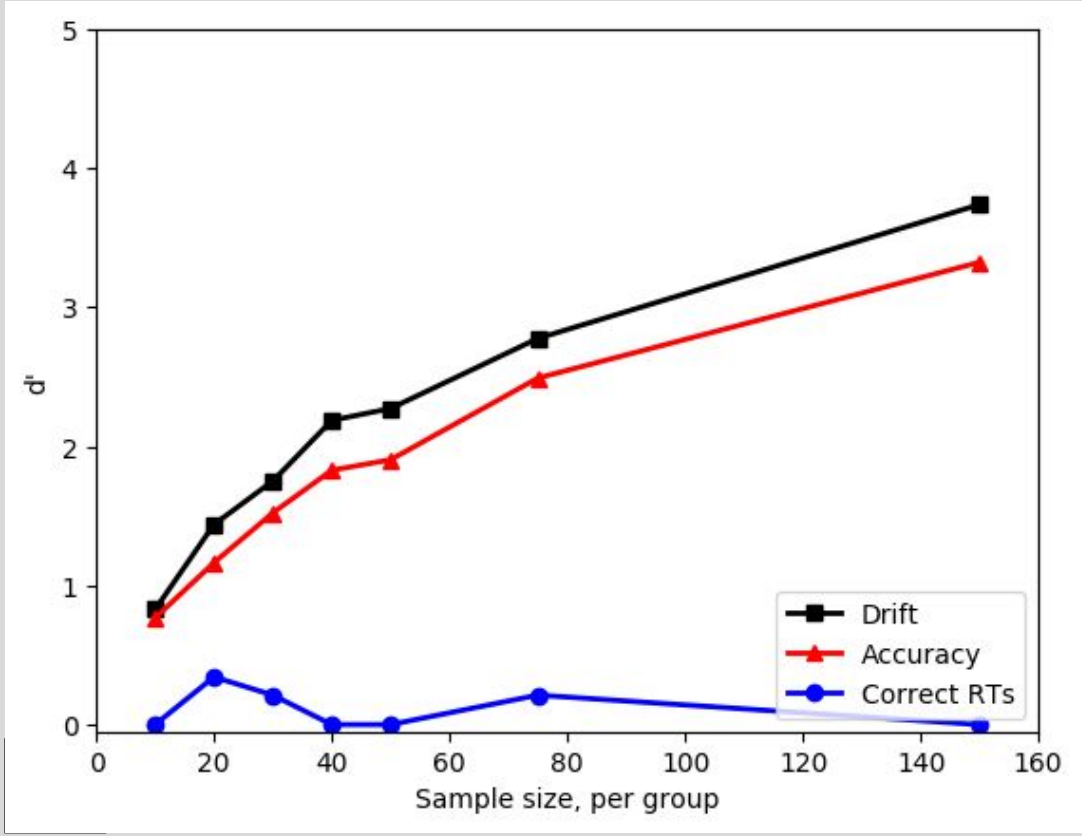
Parameters:
drift: 1 vs 1.1
boundary: 2 vs 2.1
intersubj var = 0.05
trials/ppt = 40

False Alarms

With Sato



Parameters:
drift: 1 vs 1.1
boundary: 2 vs 2.1
intersubj var = 0.05
trials/ppt = 40

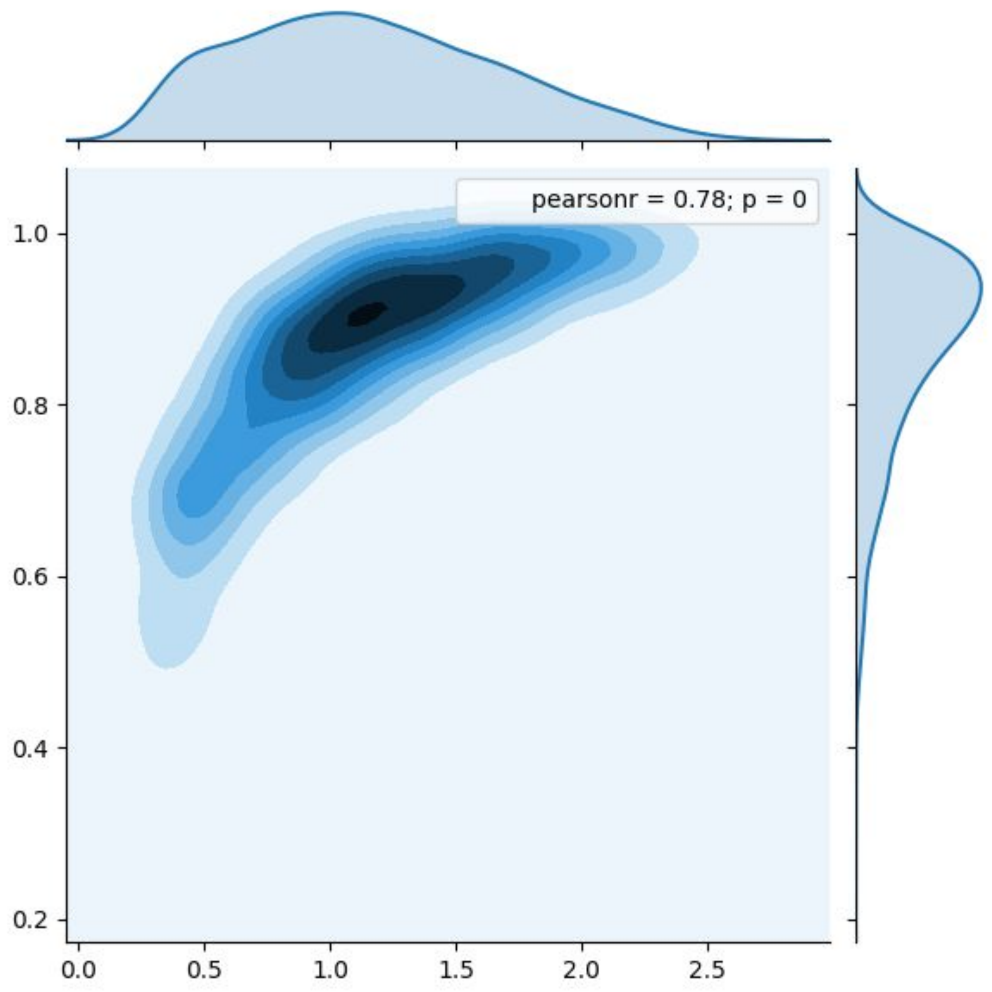


False Alarms

With Sato

Parameters:
drift: 1 vs 1.1
boundary: 2 vs 2.1
intersubj var = 0.05
trials/ppt = 40

Visualising the Speed-Accuracy trade-off


pearsonr = 0.78; p = 0
1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5

Explanatory material

Simulation strategy, part 1

1. Simulate n_e experiments where:

    Two groups, A & B, by drift and boundary parameters which are either the same or different:

        if different drift parameters: one group has superior sensitivity

        if different boundary parameters: groups make different SATOs

    n_p participants from each perform t decision making trials:

        participant drift and boundary sampled from group parameters with variation

Simulation strategy, part 2

2. Fit DDM to simulated data from each experiment.

Test difference in recovered parameters

If true difference in drift -> inferred difference = Hit

-> inferred lack of difference = Miss

If no true difference in drift -> inferred difference = False positive

-> inferred lack of difference = CR

Does it matter which decision model you use?

Obviously, yes, in some sense but

Many decision models equivalent under certain parameterisations (Bogacz et al, 2006)

Many decision models can account for any data (= unfalsifiable) (Jones & Dzhafarov, 2014)

A blind tests finds several prominent models give same inferences (Dutilh et al, 2016)

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychological review, 113(4), 700-765

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., ... & Kupitz, C. N. (2016). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. Psychonomic bulletin & review, 1-19.

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. Psychological review, 121(1), 1-32