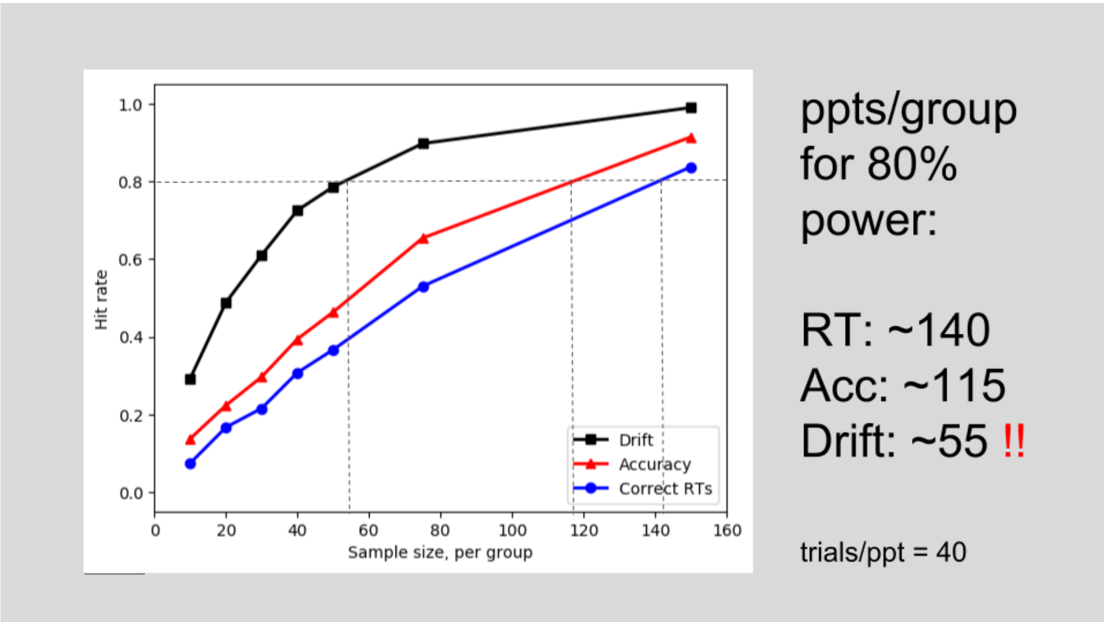
Tom Stafford
Lecturer in Psychology and
Cognitive Science
Department of Psychology
University of Sheffield
Sheffield, S1 2LT

29 October 2018

We report the results of simulations using a standard decision model and show that use of such models to fit behavioural data (accuracy and reaction times) has considerable benefit in terms of statistical power. Our analysis shows that for the simple case of a two-group test of difference in sensitivity the use of decision models can *halve* the required sample size while *still* increasing the probability of finding a true group difference and while *also* reducing the risk of a type II error due to speed-accuracy trade-off differences between groups.



This result is general to a large family of decision models which suppose the internal accumulation of noisy evidence over time to support binary choices. Such models have been the focus of intense interest over recent decades, but have not found widespread application by cognitive scientists outside of decision modelling. This paper translates the gains of research in this area so that the benefits for the design and intepretation of experiments across diverse acreas are transparent and explicit.

In the early 2000s decision modelling was the site of an important theoretical convergance. An established research programme on modelling two-choice decisions had demonstrated the suitability of a particular formulation of accumulator models – the drift diffusion model – for fitting the full distribution of response times for both correct and incorrect responses across a variety of domains and participant conditions (Ratcliff, 1978; Ratcliff & Rouder, 1998). This success was supported by neurophysiological findings showing evidence of specific neurons which acted as evidence accumulators (Gold & Shadlen, 2001; Smith & Ratcliff, 2004). These two empirical aspects were given additional meaning by demonstration that under certain parameterisations, several prominent accumulator models were equivalent to each other and equivalent to the statistically optimal method of integrating uncertain evidence over time (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Gold & Shadlen, 2002).

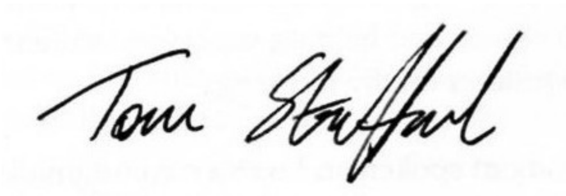
The diversity of model specifications and model fitting approachs within decision modelling may disguise the extent of consensus on the core value of accumlulator models. Additionally, a number of results point to the difficulty of distinguishing different model specifications with empirical data (Jones & Dzhafarov, 2014). This is a handicap for different modelling approaches, but a boon for experimentalists – it means that they can realise the benefits of decision modelling using any of a wide range of accumulator models. A recent multi-lab collaboration which blind-tested 17 prominent decision models against 14 different empirical data sets confirmed the inferences from all models were roughly equivalent in the majority of cases (Dutilh et al., 2016).

Given this, it is timely that the gains of decision modelling are translated to a form that is readily recognisable to non-decision modellers. Important context for this is the so-called replication crisis and the attendent methological renaissence in cognitive science. An important component of this is the attention to issues of statistcal power. Analysis of empirical work in cognitive science has show that typical statistical power is dismal (Button et al., 2013; Lovakov & Agadullina, 2017; Stanley, Carter, & Doucouliagos, 2017.; Szucs & Ioannidis; Bezeau & Graves, 2001; J. Cohen, 1962; Geuter, Qi, Welsh, Wager, & Lindquist, 2018), and this has persisted despite repeated warnings (Sedlmeier & Gigerenzer, 1989; Maxwell 2004). One reason for this is that increasing statistical power incurs costs, especially in the case of hard-to-reach populations and/or expenseive methods (such as fMRI).

The results we present will have wide appeal because they show how statistical power can be massively increased without increasing sample size or number of trials. We also show how decision models allow the principled resolution of an issue which has long dogged behavioural science, that of speed-accuracy trade-offs.

In order to maximise the impact of this work the paper is accompanied by the code for running the simulations and an interactive online data explorer, which allows the experimentalists to see the exact gains for their particular experiment parameters, in terms of statistical power and/or reduced required sample size.

Yours,



Tom Stafford

t.stafford@shef.ac.uk
0114 2226620

On behalf of

Angelo Pirrone, School of Psychological and Cognitive Sciences, Peking University
Mike Croucher, Research Computing, University of Leeds
Anna Krystalli, Research Software Engineering, University of Sheffield

References

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.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience, 14 (5), 365-376

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. The Journal of Abnormal and Social Psychology, 65 (3), 145-153.

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

Geuter, S., Qi, G., Welsh, R. C., Wager, T. D., & Lindquist, M. A. (2018). Effect size and power in fMRI group analysis. bioRxiv, 295048

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. Neuron, 36 (2), 299–308.

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

Lovakov, A., & Agadullina, E. (2017, November). Empirically derived guidelines for interpreting effect size in social psychology. PsyArXiv. doi:10.17605/OSF.IO/2EPC4

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. Psychological Methods, 9 (2), 147.

Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85 (2), 59–108.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. Psychological Science, 9 (5), 347–356.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? Psychological Bulletin, 105 (2), 309.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. Trends in Neurosciences, 27 (3), 161–168.

Stanley, T., Carter, E. C., & Doucouliagos, H. (2017). What meta-analyses reveal about the replicability of psychological research. Deakin Laboratory for the Meta-Analysis of Research. Retrieved from http://www.deakin.edu.au/__data/assets/pdf_file/0007/1198456/WhatMeta-AnalysesReveal_WP.pdf

Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLoS Biology, 15 (3), e2000797