Quantifying effect sizes in implicit learning tasks; the role of some stuff

A Preprint

Kelly G. Garner *
School of Psychology
University of Birmingham
Edgbaston, UK, B13 2TT
getkellygarner@gmail.com

Christopher R. Nolan
School of Psychology
University of New South Wales
Sydney, Australia
cnolan@cn.id.au

Abbey Nydham
School of Psychology
University of Queensland
St. Lucia, Australia, 4072

Zoie Nott
School of Psychology
University of Queensland
St. Lucia, Australia, 4072

Howard Bowman
School of Psychology
University of Birmingham
Edgbaston, Birmingham, B13 2TT
h.bowman@bham.ac.uk

Paul E. Dux
School of Psychology
University of Queensland
St. Lucia, Australia, 4072
paul.e.dux@gmail.com

June 10, 2021

Abstract

Enter the text of your abstract here.

***K*eywords** blah · blee · bloo · these are optional and can be removed

1 Introduction

~100 words per paragraph

The brain is the most complex organism known to humans, yet decision making regarding theory for its function tends to be made on binary (i.e. pass or fail) terms, at least in the experimental psychological sciences. Specifically, theories often propose experimental tests for the presence or absence of given effects, rather than quantifying the extent to which an effect should be observed, i.e. the anticipated effect size. The latter prediction is more risky, and therefore constitutes a more desirable prediction for theory testing [insert Popper reference]. In fact, it seems unlikely that such pass/fail decision-making will be sufficient to disentangle the myriad functional systems that the brain has developed over millions of years of evolution.

For example, in the study of EF: [AB: theory of p/f vs Ragnaroc?], MT costs - .

Likewise, in the study of IL: VSL, CC, SRT

To promote quantised theories in experimental psychology, one extra piece of pertinent information that is informative to theory development is - we normally do x (insert Cummings refs), but it is missing y.

This is also useful for experimental development. i.e. The other important thing is that to ensure that we provide sufficiently precise information, so that this can be used. To do that we need to perform power calculations, There are at least 3 ways to do this: have a sufficiently precise theory that quantifies the effect size of interest, arbitrarily assume an effect size of theoretical interest, but unmotivated, or create an estimate by sampling the field. The problem with the latter is that given the current sample sizes typically employed, we have no idea if that estimate is precise and should be used.

A method to determine the precision of effect

*Corresponding author

For example, theories in the experimental psychological sciences have tended only to predict the presence or absence, rather than the extent of a given phenomena. For example [example from implicit learning]. However, clear to see that if we are to understand how such a complex system processes sensory information, coordinates tasks and acquires new behavioural repertoires, a precise mapping between theory and outcome is going to be necessary. [get some Vehicles notions in that previous sentence]. Therefore important to start thinking about the size of effects - however, current state of field, very little knowledge about anticipated effect sizes due to x, y, & z.

Using a large dataset we will address this gap. Data on x-tasks. We will apply a simulation analysis to determine x, y and z.

Moreover, we will address two pertinrent analytical gaps: i) A further development is the recent use of linear mixed effects models, and the recommendation that we use them instead of ii) it is common to use t-test on accuracies against chance but Allefeld (VSL).

PUT IN INTRO Additionally, given the documented advantages of linear mixed effects models (LME) over repeated-measures ANOVA (Muth et al. 2016; Bagiella, Sloan, and Heitjan 2000; McCulloch 2005), and that only a conceptual proxy of η_p^2 is computable from these models (Brysbaert and Stevens 2018; **???**), and c) there exists no data that we know of that quantifies to what extent we can expect comparable outcomes between both methods, we (where relevant) opted to apply both the commonly used statistical model, and a LME model to each k sample

2 Methods

I'll insert a summary sentence here during write up.

See Section 2.

2.1 Participants

The current study uses a data set collected for a previous pre-registered project examining the relationship between executive function and implicit learning. This data set contains performance measures from N = 313 participants. Participants were undergraduate students, aged 18 to 35 years old (mean = 20.14 yrs, sd = 3.46). Of the total sample, 208 reported being female sex, and 269 reported being right handed. Participants received course credits as compensation. All procedures were approved by the University of Queensland Human Research Ethics Committee and adhered to the National Statement on Ethical Conduct in Human Research.

2.2 Apparatus

Experimental procedures were run on an Apple Mac Minicomputer (OS X Late 2014, 2.8 GHz Intel Core i5) with custom code using the Psychophysics toolbox (v3.0.14) (Brainard 1997; Pelli 1997) in Matlab v2015b. Participants completed 5 tasks; Attentional Blink (AB), Dual Task (DT), Contextual Cueing (CC), Serial Response Task (SRT), and Visual Statistical Learning (VSL). Task order was randomised for each participant, apart from the VSL task, which was presented last. This was because the recognition component of the task may have allowed participants to infer that other tasks were also assessing implicit learning.

2.3 Procedures

Across all tasks, participants sat approximately 57 cm from the monitor. An overview of the task procedures is presented in Figure **??**. Further details regarding the task protocols are presented within each section below. In the interest of reducing working memory load, we provide an overview of the simulation procedures, before detailing the specific procedural and statistical methods for each task.

NEXT STEP: check this: `https://bookdown.org/yihui/bookdown/figures.html`

All the data and code used for the current analysis are available online. The analysis of the data from each task followed two steps. First, to ascertain that we observed the typical findings for each of the paradigms, we applied the conventional statistical model for that to the full dataset (N=313). The details of each analysis are presented below. Next, we implemented a simulation procedure to determine the effect size and p-values that would be attained over many experiments conducted at multiple levels of sample size.

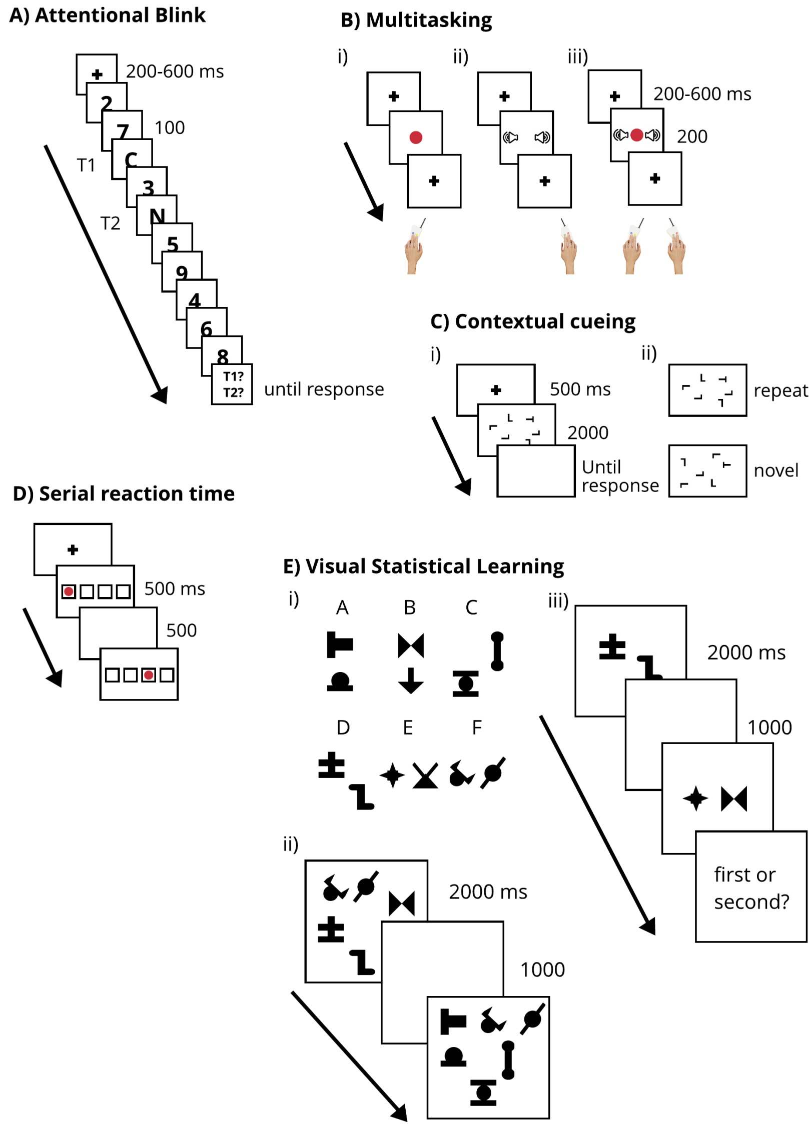


Figure 1: Task battery. A) Attentional Blink Paradigm (AB). Participants report the two letter targets from the rapid serial visual presentation of numbers and letters, B) Multitasking Paradigm (MT). Participants make a discriminate the colour of a disc, a complex tone, or both C) Contextual Cueing Paradigm (CC). i) Participants perform an inefficient visual search task. ii) Unknown to participants, half of the search arrays are repeated throughout the course of the experiment. D) Serial reaction time task (SRT). Participants respond to one of four stimuli, each mapped to a spatially-compatible button press. Unknown to participants, for half of the blocks the stimulus follows a repeating sequence. E) Visual Statistical Learning Paradigm (VSL): i) 12 shapes are grouped into 6 base pairs. ii) Learning: three of the six pairs are presented as an array, this is repeated as participants passively view the displays. iii) Test: participants are presented with a base pair, and a novel pair formed from a recombination of the 12 shapes, and is asked which of the two pairs they have seen previously.

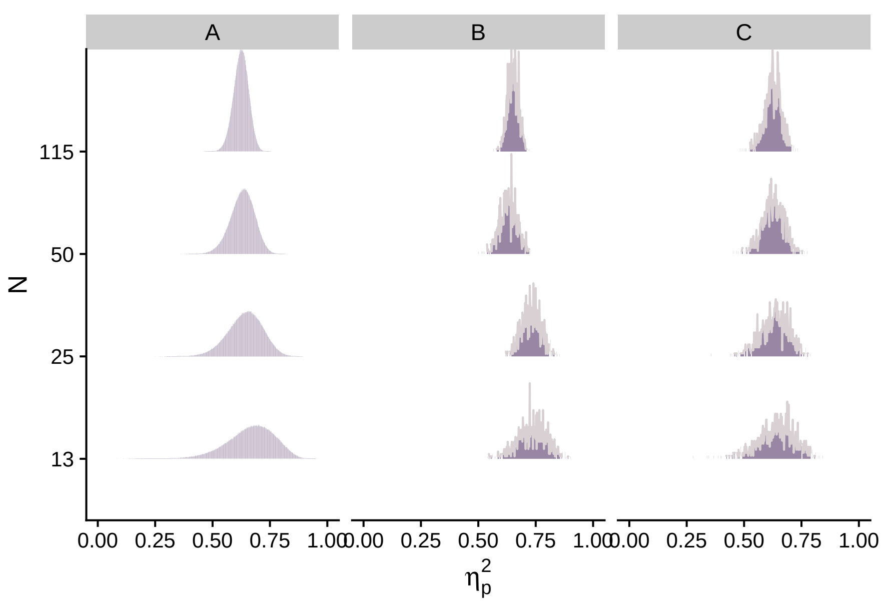
Figure 2: Comparison between sampling procedures for the AB task data fit with a repeated measures ANOVA for 4 levels of N; densities of observed effect sizes for the A) two-step sampling procedure j=1000, k=1000, the B) two-step sampling procedure, j=1, k=1000, and C) the one-step procedure, $k = 1000$

2.3.1 Sampling procedure

For each task, we sampled across 20 different sample sizes (Ns), defined on a logarithmic interval between N=13 and N=313. We opted for a logarithmic interval given the decreasing information gained at higher N values. To simulate k experiments at each of our chosen N, we developed a sampling procedure that sought to leverage information from across the whole dataset while also protecting against any reductions in effect size variablility that may be attributable to saturation as N approaches the maximum (N_{max}=313). Specifically, it could be that as N approaches 313, the overlap of participants between subsamples may be greater than when N equals a lower number such as 13. It follows then that any decreasing variability in effect size estimates at higher Ns could be due to the decrease in variability of the subsamples, rather than the improved estimate of the population variance should come with a larger N.

To protect against this possibility we applied the following procedure; for each level of N ($N_1, N_2, ...N_{20}$), we first selected a subsample from the total dataset *without* replacement, e.g. N of 13 unique samples from the total N=313. We refer to this from now on as the *parent subsample*. From this parent subsample, we sampled $k = 1000$ times *with* replacement - e.g. N=13 sampled from N=13. These shall now be referred to as the *child subsamples*. The relevant analysis was then applied to each of the child subsamples. Sampling with replacement ensured that the child subsamples carried the Markov property. Given that this procedure reduces the heterogeneity of the parent sample (for example, in the case of N=13, all the child subsamples are derived from only 13 unique observations), we then repeated the entire process over $j = 1000$ iterations. Therefore the presented data reflects $j * k$ simulated experiments for each level of N. We refer to this now as the two-step sampling procedure: $j = 1000$, $k = 1000$. It is worth noting that we compared this procedure to one where we performed the two-step sampling procedure with $j = 1$, and to a one-step sampling procedure where we sampled N with replacement from the entire dataset, i.e. k=1000 and $j = 0$. Outcomes were comparable between the sampling procedures, with the two-step procedure ($j = 1000$) offering better resolution of the resulting densities (see Figure 2 for a representative example).

```
## ‘summarise()‘ regrouping output by ’samp’, ’mod’ (override with ‘.groups‘ argument)
```



$$\xi_{ij}(t) = P(x_t = i, x_{t+1} = j|y, v, w; \theta) = \frac{\alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}$$

2.3.2 Headings: third level

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Paragraph Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

3 Examples of citations, figures, tables, references

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui. some text (**???**; **???**) and see (**???**).

The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

> Hasselmo, et al. (1995) investigated...

`https://www.ctan.org/pkg/booktabs`

3.1 Figures

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi. See Figure 3. Here is how you add footnotes. [^Sample of the first footnote.]

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetuer eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

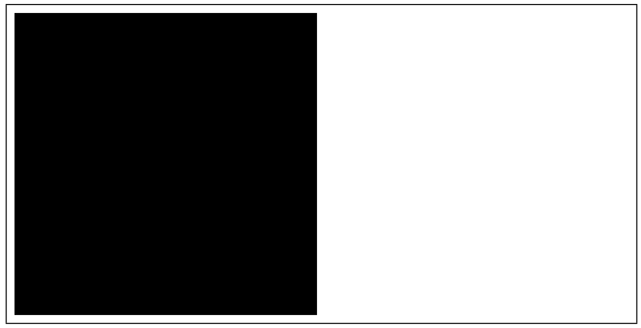
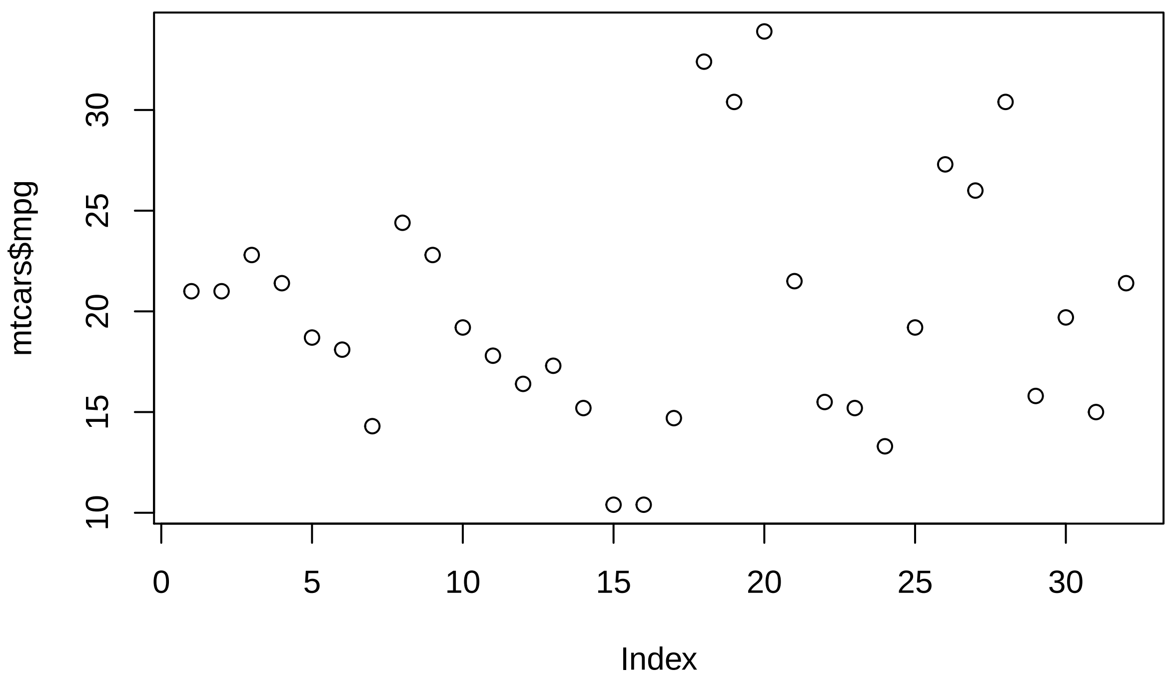


Figure 3: Sample figure caption.

Table 1: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6



3.2 Tables

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

See awesome Table~1.

3.3 Lists

- Lorem ipsum dolor sit amet
- consectetur adipiscing elit.
- Aliquam dignissim blandit est, in dictum tortor gravida eget. In ac rutrum magna.

Bagiella, Emilia, Richard P. Sloan, and Daniel F. Heitjan. 2000. "Mixed-Effects Models in Psychophysiology." *Psychophysiology* 37 (1): 13–20. `https://doi.org/https://doi.org/10.1111/1469-8986.3710013`.

Brainard, D. H. 1997. "The Psychophysics Toolbox." *Spatial Vision* 10 (4): 433–36.

Brysbaert, Marc, and Michaël Stevens. 2018. "Power Analysis and Effect Size in Mixed Effects Models: A Tutorial." *Journal of Cognition* 1 (1): 9. `https://doi.org/10.5334/joc.10`.

McCulloch, Charles E. 2005. "Repeated Measures ANOVA, R.I.P.?" *CHANCE* 18 (3): 29–33. `https://doi.org/10.1080/09332480.2005.10722732`.

Muth, Chelsea, Karen L. Bales, Katie Hinde, Nicole Maninger, Sally P. Mendoza, and Emilio Ferrer. 2016. "Alternative Models for Small Samples in Psychological Research: Applying Linear Mixed Effects Models and Generalized Estimating Equations to Repeated Measures Data." *Educational and Psychological Measurement* 76 (1): 64–87. `https://doi.org/10.1177/0013164415580432`.

Pelli, Denis G. 1997. "The VideoToolbox Software for Visual Psychophysics: Transforming Numbers into Movies." *Spatial Vision* 10 (4): 437–42. `https://doi.org/10.1163/156856897X00366`.