Twitter Thread by James Bartlett

James Bartlett
@JamesEBartlett

It was great to talk about reproducible workflows for @riotscienceclub @riotscience_wlv. You can watch the recording below, but if you don't want to listen to me talk for 40 minutes, I thought I would summarise my talk in a thread:

> Thank you again @JamesEBartlett for a fantastic talk (with a really nice personal touch) on reproducible workflows!
>
> Thanks especially for the co-leads @IMLahart for co-hosting and @DrManiBhogal for nabbing James!
>
> Slides: https://t.co/CNqxzOhch1
> Video: https://t.co/YjHEHuRJlz
>
> — RIOT Science Club Wolverhampton (@riotscience_wlv) February 16, 2021

My inspiration was making open science accessible. I wanted to outline the mistakes I've made along the way so people would feel empowered to give it a go. Increased accountability is seen as a barrier to adopting open science practices as an ECR https://t.co/bNK6poNpzx

It also comes across as all or nothing. You are either fully open science or your research won't get anywhere. However, that can be quite intimidating, so I wanted to emphasise this incremental approach to adapting your workflow https://t.co/HS3dx5qJL3

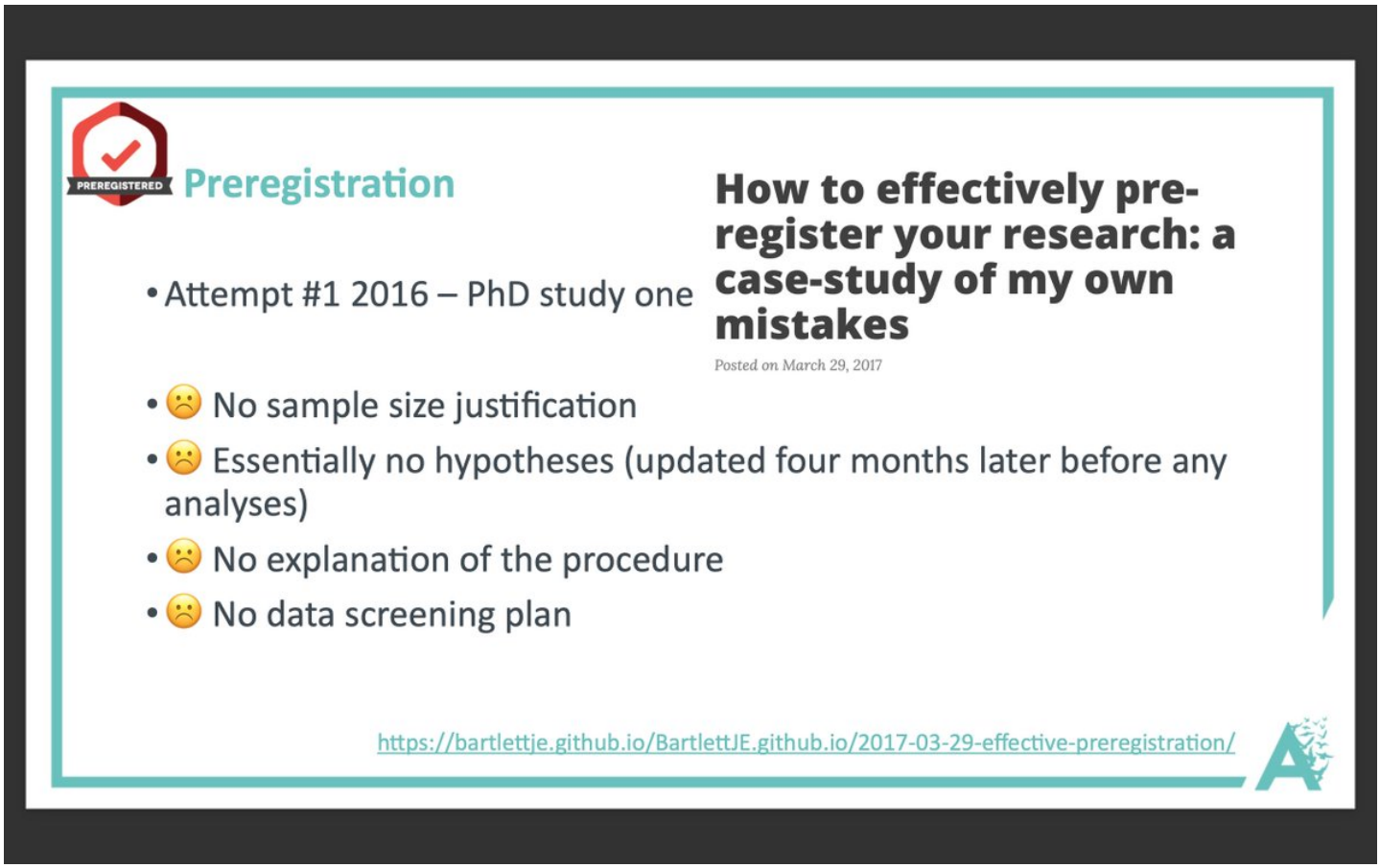
There are two sides to why you should work towards reproducibility. The first is communal. It's going to help the field if you or someone else can reproduce your whole pipeline.



There is also the selfish element of it's just going to help you do your work. If you can't remember what your work means after a lunch break, you're not going to remember months or years down the line https://t.co/iRSXISI5NE

> Why are lunch breaks important for #code?
>
> If you can't remember what your variable names refer to after lunch, you sure as hell won't remember in 3 months.
>
> — Dr Rebecca Hirst (@HirstRj) February 11, 2021

Pre-registration is super helpful, but it takes time to learn how to do it effectively. I had no support for my first attempt and it was awful. https://t.co/o0QgrtBFfJ



But I learnt from my mistakes and eventually got to a point where I'm happy with my approach.

Preregistration

- Attempt #3 2019 – PhD study three

- 😃 Reproducible simulated power analysis in R Markdown
- 😃 Specific hypotheses
- 😃 Exclusion criteria and analysis plan outlined using simulated data

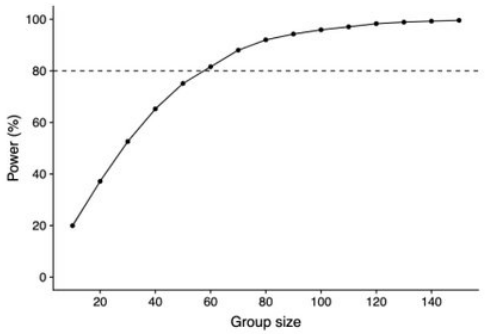


Figure 3: Power curve for the main effect of smoking group on attentional bias with increasing group size. Dashed line indicates 80% power.

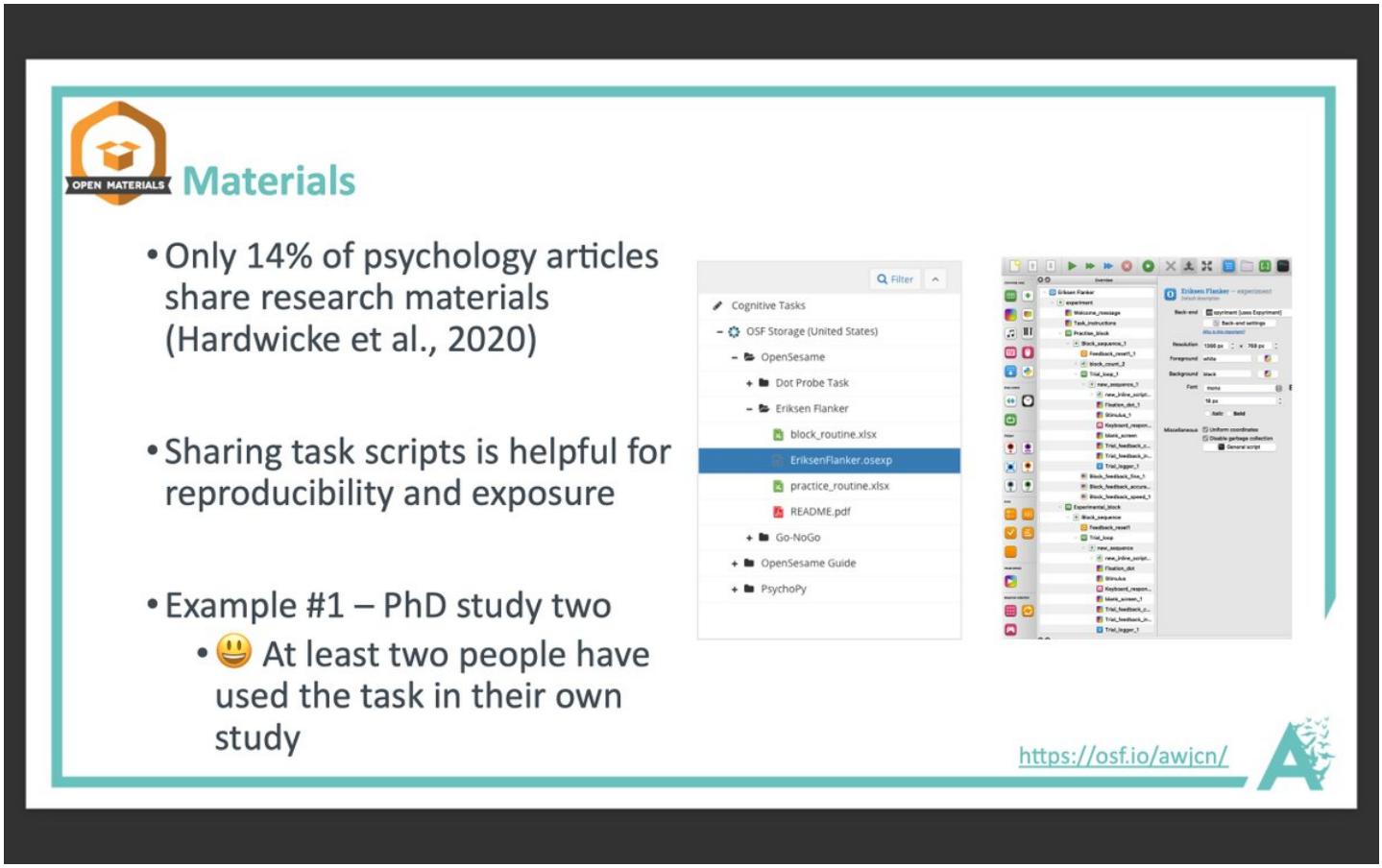
https://osf.io/hjk5u/

Some aspects of open science can't be added at the end. If you want to share your data, it has to be planned from the ethics stage and your participants have to consent to it. Just remember not all data can be shared openly for ethical reasons.

Ethics

- Example #1 – PhD study one
  - ☹ Open data not mentioned on the consent forms
- Example #2 – PhD study two
  - 😃 Open data clearly outlined on the consent forms
- Example #3 – PhD study three
  - 😃 Open data clearly outlined on the consent forms

Sharing materials is one of the easiest things you can do, but very few researchers do it. You can share your stimuli (providing there's no copyright issues) and your task scripts. If you use @GorillaPsyc, there is an open materials function making it super easy.

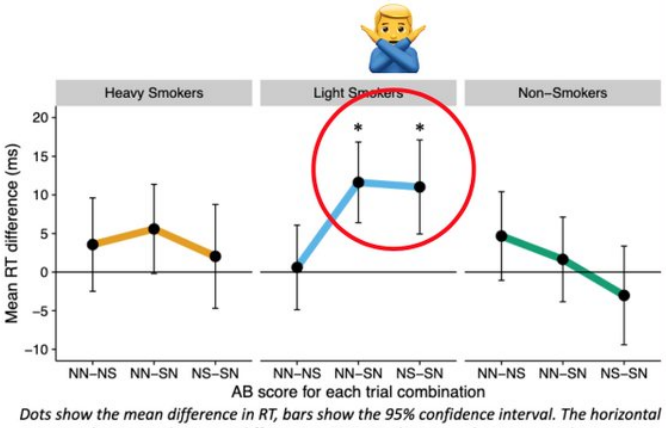


Double check your analysis pipeline, especially when the results support your hypotheses. I caught an error in my first PhD study before it was too late, but it could have been super embarrassing.

Analysis

Example #1 (PhD study one)

- 😬 Originally processed the response time data in Excel and copied into SPSS
- Months later, I checked the analyses using R and realised I mixed up two columns
- 🙃 All the effects disappeared when I corrected the processing pipeline



Dots show the mean difference in RT, bars show the 95% confidence interval. The horizontal bar at 0 indicates no difference in RT. Note the unequal y axis around 0.

You don't have to go the full R Markdown manuscript route, but I recommend at least scripting any data processing. If you do make a mistake, at least that error is reproducible. You can also use intermediaries like @tidystats to embed results from R into your Word file.

Analysis

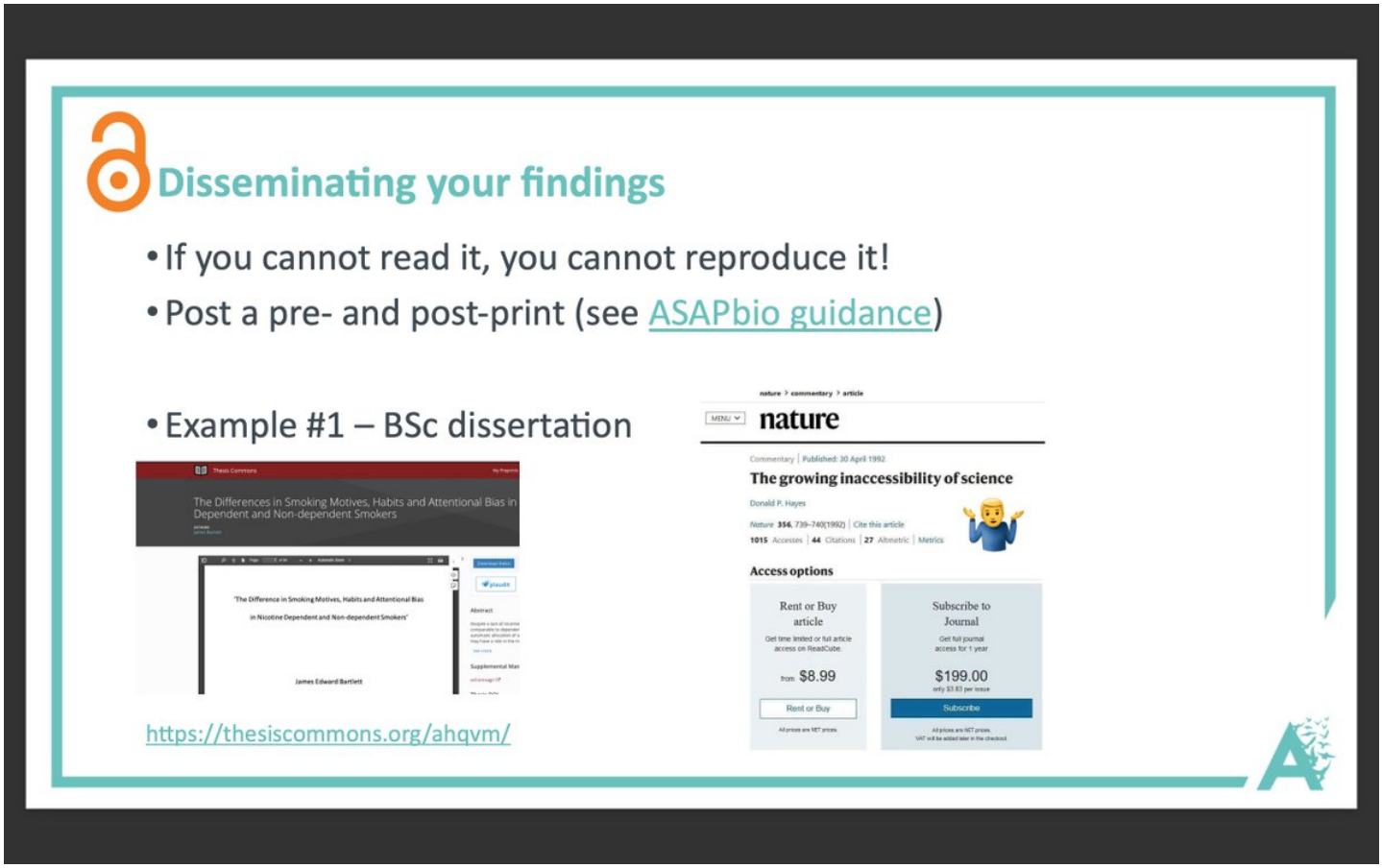
Example #3 PhD study three

- 😃 Open data outlined in consent forms
- 😃 Reproducible manuscript written in R Markdown
- 😃 Includes the whole analysis pipeline from raw data to graphs



https://osf.io/6vzba/

Finally, if you can't read the article in the first place, you can't reproduce it. Most journals allow you to post pre/post-prints and you can upload your dissertation/thesis to Thesis commons. Just double check it's permitted by your target journal: https://t.co/iIIw1VDHdC



Hopefully, this seminar shows people fuck up from time to time. It's about learning from those mistakes and doing what you can to minimise them in future. Try making your workflow more reproducible one step at a time. It'll be worth it in the long-run for you and the community.