Gathering data online
Tools and platforms

Stephanie Droop
stephanie.droop@ed.ac.uk
Intro

- My name is Stephanie Droop
- I'm a second year PhD student in the NLP CDT at the University of Edinburgh
- Psychology background
- Research interests focus on cognitive modelling
- I'm going to talk about some of the tools and platforms I've used to collect behavioural data
Overview

A big part of psychology or NLP research is collecting data online in two different ways:

1. Large data sets which need annotated. Time consuming for one person. Use crowdsourcing platforms or annotation tools
2. Behavioural experiments to better understand humans

Today we’re talking mainly about the second of these but first we’ll quickly mention annotation.
Annotation data

- Given some set of data X, you want labels Y
  - For example, you have photos you want labeled to say if the person is young or old
  - Or tweets that you want to label as sarcastic or serious
- Probably going to use it as training data for some model
- Or use it as ground truth demographic data for statistical analysis
- Generally high volume
- Data to be labeled is "real" (potentially messy)
- Don't care too much about who is doing the labelling (usually)
  - But you want quality controls
Tools for annotation: Appen - benefits

www.appen.com

- Templates for common use cases (text or image labelling, judgement, transcription etc.)
- Easy to customise using HTML-like template language
- Data is imported by CSV upload
- Good quality control options to filter out bad annotators
- Great options for aggregating/filtering your final data
- Great for getting multiple annotations per item
- Very fast turnaround
- Deals with all issues of payment and so on
Tools for annotation: Appen - issues

www.appen.com

- No pre-filtering of annotators based on demographics
- Terrible for complex tasks, due to oddly limited template language
- Even if you think your task is very simple, some people will get it wrong because they didn't read your instructions
- Doesn't matter how good your written instructions are
Tools for annotation: Prodigy

- More focused on natural language processing data gathering - POS tagging, NER tagging
- But also images
- From people who make spacy
- Great interface
- Use ML to speed up labelling
- Survey paper:

  An extensive review of tools for manual annotation of documents (2021)
Collecting behavioural data online

- Given some stimulus, how does a person react?
- Aim is to say something about people
- Probably going to do statistical analysis on results
- Or try to model these results using cognitive modelling to end up with a) better interpretation of patterns of behaviour, b) plausible process of what may be going on in the mind when people do a task
- Generally lower volume (depending on design/power)
- Data used is generally highly controlled, possibly artificial
- Care very much about who is taking part
  - And still want quality controls
Two parts

Your study is in two parts:

1. Make the actual experiment using code or interface, eg. jspsych [jspsych.org](http://jspsych.org)
2. Issue it to participants online using a platform which handles recruitment, data and payment, eg. Prolific [prolific.co](http://prolific.co)
jsPsych

[jspych.org](jspych.org)

- JavaScript library with plugins for a variety of experimental tasks
- Support for text, image, audio and video stimuli
- You need to manually create your stimuli and balance them across any conditions you have
  - Generally, use Python or whatever to generate blocks of stimuli
  - Keep them in the same folder as the experiment script
  - Run copies of the same experiment for each block
- Very easy to make changes, as stimuli and experiment code are separate
- See example shown in class
“Online Experiments for Language Scientists” course

Some of you may be doing this course by Kenny Smith in PPLS

Even if not, you can access the materials.  
[https://kennysmithed.github.io/oels2022/](https://kennysmithed.github.io/oels2022/)

I found this course great and gives a lot to copy / emulate.

BUT it skipped over the basics of javascript, the ‘filler’ statements or functions in between jsPsych nodes.
You will also need…

- General javascript, html and CSS. Try this course https://softdev.ppls.ed.ac.uk/online_experiments/
- (Downside: it can take a newbie a few weeks to get an experiment working, and even then it may look plain)
- A web server to host on: ask Frank or your supervisor
- Sound experimental design, stimuli, research question, etc!
- Canny attention checks
- By the end of this stage you should have a website that if you go to, it starts your experiment as if you are a participant
- Now you need a place to host your experiment…
Experiment host sites

- MTurk - full of bots, traditionally not paid enough
- Testable [https://www.testable.org/](https://www.testable.org/) - use this if you want to avoid coding - but the trial file can be fiddly and you may wish you’d invested that time in learning javascript
- Prolific - the current favourite if you have already coded your experiment using eg. jsPsych
Prolific

prolific.co

● Only a front-end for recruitment
  ○ Your actual experiment must be hosted elsewhere!
● Very fine-grained control over participants
  ○ Only limit is how many people on the platform meet your requirements
● Participants tend to be higher quality, better at reading instructions
● Deals with all aspects of payment
  ○ Minimum payment of ~£5/hour pro-rata
  ○ Can refuse to pay participants if they fail some attention check, take too long, or so on
  ○ In this case you need to manually open each ppt’s data and check, then press approve
Prolific (2)

prolific.co

- Can be fiddly if you want to run several blocks
  - There is a setting for excluding the IDs of people who have done it already
- To actually get funds loaded: email Sally with cost breakdown
  - E.g. “200 ppts for 20 mins each @ £2.50, plus Prolific’s 30% cut = …”
  - Then [probably?] you’ll generate an invoice on prolific site then copy on to Nikolaos
  - Shouldn’t need to pay yourself
Summary

- Appen is good for large scale simple data labelling
  - But I would never run actual experiments on it
- Prodigy is great for annotating NLP data
  - But not sure how you can scale it beyond a few people
- Prolific is good for recruiting specific types of people, very conscientious
- jsPsych is a good toolkit for basic behavioural experiments
- Testable is a good code-free option if this is one-off
- If you want to collect data more than once throughout your career, learn js
End

So that's an overview of the tools and platforms we commonly use for research.

For annotation, ask on slack for other advice.

Thanks for your attention and I hope your research goes well!