5 Things to Consider When Collecting Research Data Online

Sydelle de Souza
Hi, I'm Syd!

2nd Year PhD Student (CDT in NLP)
(sydelle.desouza@ed.ac.uk)

- Academic Background:
  - BA Theoretical/Portuguese Linguistics
  - MSc Second Language Acquisition
  - MSc Psychological Research
- Now: Computational Cognitive Science
- Research Interests:
  - Language Acquisition & Processing
  - Compositionality & Semi-Productivity
  - Analogical Reasoning
- Research Experience
  - Experimental Morphology
  - Experimental Psycholinguistics
  - (Some) Machine Learning
WHY COLLECT DATA ONLINE?
PROS

- Faster data collection
- Larger samples
- Cost effective
- Diverse populations
CONS

COMPREHENSION

ATTENTION

HONESTY

RELIABILITY

DATA QUALITY

Depends on:
- Comprehension
- Completeness
- Accuracy
- Representation
- Validity

Is the data fit for its intended purpose? Can you extract usable & valuable insights?
WHAT CAN WE DO TO IMPROVE DATA QUALITY?
#1 Use the right tool for the right job
WHAT ARE YOU COLLECTING?
WHAT ARE YOU COLLECTING?

ANNOTATIONS

SURVEYS
WHAT ARE YOU COLLECTING?

- Annotations
- Behavioral Data
- Surveys
WHAT ARE YOU COLLECTING?

ANNOTATIONS

BEHAVIORAL DATA

PRECISION & CONTROL

SURVEYS

MODERATE

HIGH
WHAT ARE YOU COLLECTING?

ANNOTATIONS

prodigy
tappen

BEHAVIORAL DATA

SurveyMonkey
qualtrics

SURVEYS

Typeform

PRECISION & CONTROL

jpspsych
GORILLA

MODERATE

Testable

millisecond

HIGH

PsychoPy
#2 WELL DESIGNED IS HALF DONE!
BAD DESIGN LEAVES PARTICIPANTS...

CONFUSED  BORED  EXASPERATED  OUTRAGED  ENRAGED
YOU WANT YOUR PARTICIPANTS TO BE...

RELAXED  ENGAGED  MOTIVATED
FOLLOW GRICE'S MAXIMS

1. Concise, Clear Instructions
2. Orderly Task Flow
3. Keep It Short & Relevant
4. Keep It Simple

#3 PILOT TWICE, RUN ONCE!
ROBUSTNESS CHECK

- Before deployment:
  - Work out kinks in your design
  - Spelling & grammar checks
  - Inconsistencies in presentation
- Use it to build your analysis pipeline
  - Add/remove variables
- Pilot with experts AND non-experts
- Get as much feedback as possible
  - IMPLEMENT IT!
- Saves time & £££
# 4 KNOW THY PARTICIPANTS
PRE-SCREENING

- Balanced and representative sample
- Ensure that your participants fit the profile of your target audience
- Pre-determine the demographic variables of interest. For example,
  - Age
  - Gender identity
  - First language
  - Employment status
  - Social media use
Beware professional survey takers

- Non-naive participants are problematic!
  - Introduce biases
  - Reduce effect sizes
  - Provide poor responses
- Negatively impacts data quality


Watch: https://www.pbs.org/newshour/science/inside-amazons-hidden-science-factory
# 5 YOU GET WHAT YOU PAY FOR
COMPENSATION

- Participants expect to be paid for their time!
  - Choose the right incentive
  - Consider a performance bonus
- Is your task unusually long, mentally taxing, or uses stimuli that cause emotional distress?
  - If yes, then increase the reward!
- An attractive reward == increased motivation
  - Reduces drop-outs
- Poorly thought-out incentives can skew decision making

GOLDEN RULE:
DO UNTO OTHERS
FOR ANNOTATION TASKS:
FOR EVERYTHING ELSE:
SYD DOESNAE RECOMMEND:

amazon mechanical turk

RESOURCES

- Collecting behavioral data online (APA)
- So, you want to run an online experiment?
- nodeGame Online Real-Time Synchronous Experiments
- Survey data quality: The 4 factors that matter most to researchers
- Reddit discussion on survey platforms
- How to Make $50 a Day on MTurk: How is everyone doing it?
- How to increase participant attention in online experiments
- Pay Fairness
OOK!