From good to better

Skills for executing a transparent research project

Candice C. Morey and Richard D. Morey
Canadian Society for Brain, Behaviour, and Cognitive Science / Experimental Psychology Society

Open Science on the Rock!

From Good to Better:

Skills for Executing a Transparent Research Project

Featuring

Candice Morey, Stephen Lindsay, Shannon Cobb & Alexander Etz

I. Motivation

Adopting transparent practices

- Scientists endorse openness, but most don't prioritize it
- Perceived to be a lot of work
- Rewards fuzzy:
  - Idealistic only?
  - Maybe important later, but not vital now?

Strategic concerns

- Won't it slow me down?

Strategic concerns

- Why should someone else benefit from my work?

Strategic concerns

- What if I fail?

Science is a collaborative effort

You are working together to achieve something bigger.

Why you need transparency

Your project

- Will not go as planned
- Something unexpected will happen
- Some of your studies will not work as intended
- Some of your chapters will not be accepted on the day you submit
- You may start working on something else days after submitting

Scenario: You're finished with your project!

Scenario: A road-block to publication!

You will need your lab's ongoing support to make the most of your work.

Why PIs need in-lab transparency

Journal of Memory and Language 69 (2013) 324–348

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Asymmetric cross-domain interference between two working memory tasks: Implications for models of working memory

Candice C. Morey [a,*], Richard D. Morey [b], Madeleine van der Reijden [a] (MSc, 2010), Margot Holweg [a] (MSc, 2011)

[a] Experimental Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712TS Groningen, The Netherlands
[b] Psychometrics and Statistics, University of Groningen, Grote Kruisstraat 2/1, 9712TS Groningen, The Netherlands

Why PIs need in-lab transparency

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000280

Spatial Sequences, but Not Verbal Sequences, Are Vulnerable to General Interference During Retention in Working Memory

MSc 2014

Candice C. Morey and Monica D. Miron
University of Edinburgh

Why PIs need in-lab transparency

Volunteer RAs 2015-2016 Volunteer RAs 2012-2013

We thank Tuulia Torppa, Josephine Steeghs, Sabine Kästner, and Leire Martin Mendez for assistance with data collection related to this project. Preliminary data influencing the experiments reported here were collected at the Heymans Institute for Psychological Research at the Rijksuniversiteit Groningen. Experiment 1 formed part of Monica Miron's Master of Science dissertation at the University of Edinburgh. Data, analysis scripts, and experimental software are freely available on our Open Science Framework page (https://osf.io/s6w4b/).

Each student thinks their project is theirs, but it is always part of something bigger.

Preparing for openess is simple

- It's just documenting your work, making it easier to understand
  - So that lab members can learn from each other
  - So that you don't forget what you've done
  - So that your colleagues can help your work flourish

Preparing for openess is simple

- It's just documenting your work, making it easier to understand
  - So that lab members can learn from each other
  - So that you don't forget what you've done
  - So that your colleagues can help your work flourish
- Has the bonus of allowing you to share outside, if/when you want

We're convinced! Let's develop a system.

II. Implementing transparent research practices

Assume eventual transparency from the start

- Remember: you are sharing at least with Future You, colleagues
- Assume you/they will need to understand this project (you/they will)
- Future You? You, in 6 months? 6 years? What will you remember?

Being considerate of Future You saves time

- Blog post about archiving a data set on request 10 years after it was published.
  - Could do it, but spent a lot of time double-checking and organizing
- Future Me was inconvenienced a lot coping with switching between teaching and research in first post-training academic job.
- Worth it to have a system in place!

Developing a system

Think of the information you or your colleagues will need to write a paper:

- The actual data
- Information about how it has been processed
  - Participant exclusions?
  - Scoring procedure?
  - Experimental protocol, log?
  - Data collection software/materials
- Where this stuff is located, who can access it
- How did you ensure its quality? Can you back this up? Will you remember you did it in 6 months from now?
- How does it work? Would future you or someone else be able to figure it out?

Developing a system

Organizational conventions

- How do you store and find things? Is it:
- Hierarchical?
- Consistent?
    - Similar workflow per section?
    - Repeated structure
- Secure?
    - How much would a crash set you back?
    - Will materials be findable and openable in 6 months? 6 years?

Making it better

- The system I used as a grad student and an ECR ticked these boxes.
- When I wanted to reproduce an analysis years on, it worked.
- But there were a lot of weak spots where something could go wrong, and it was inefficient. It cost time to use this system.

We can do better.

Working in semi-public on Open Science Framework

- OSF: A tool you can use to organize, back-up, and eventually share finished components
- Sharing your finished products != working totally in public
  - Custom software, stimuli, protocols: they're finished when study is ready to be run
  - Pre-registrations: Before analyzing new data
  - Anonymized data: when the data set is closed to new members, entries
  - Analysis code and papers that are ready for scrutiny from someone

Nice features of OSF

- It's free
- Control project members, visibility
- Can get DOI assigned to project
- Free storage

Next steps:

- We're going to implement a project on OSF.
- We're going to mimic several key stages of a research project:
    - Planning
    - Starting data collection
    - Data analysis/Communication
- Note: This is a toy! A real project may be more complicated.

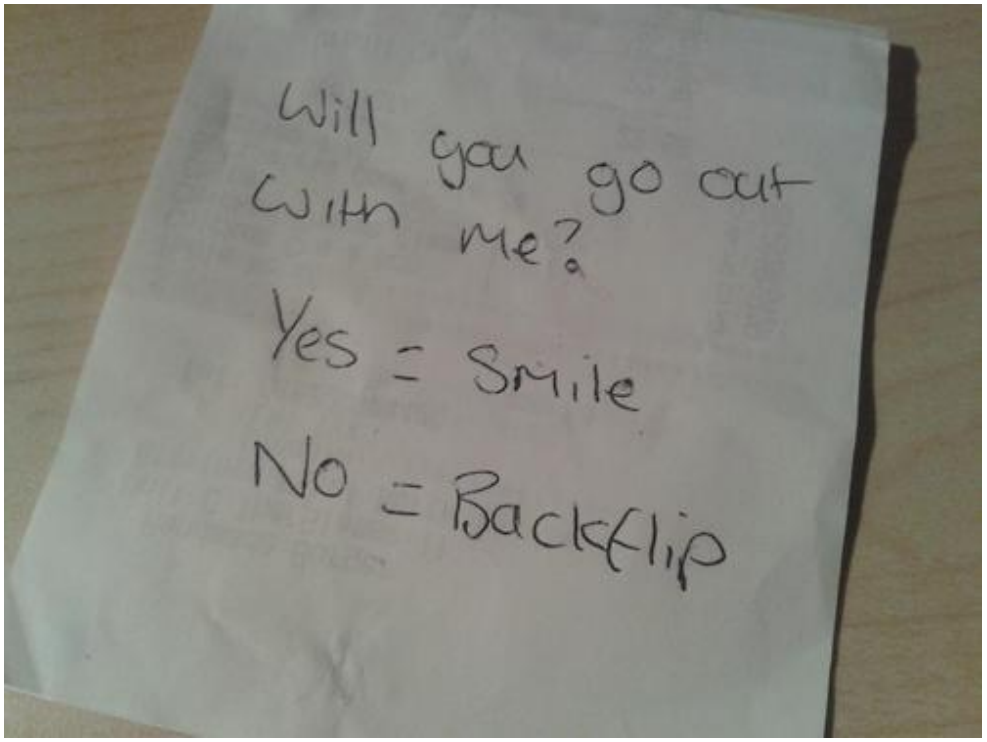
The OSF environment

https://osf.io/

- Your account: associates all of your projects with your id
- Project: Space associated with your "project"
- Project can be one study, multiple studies
- Project space should include all components: rationale, materials, data, analyses, paper
- Components may be selectively shared with specific users, or selectively made public

Our project

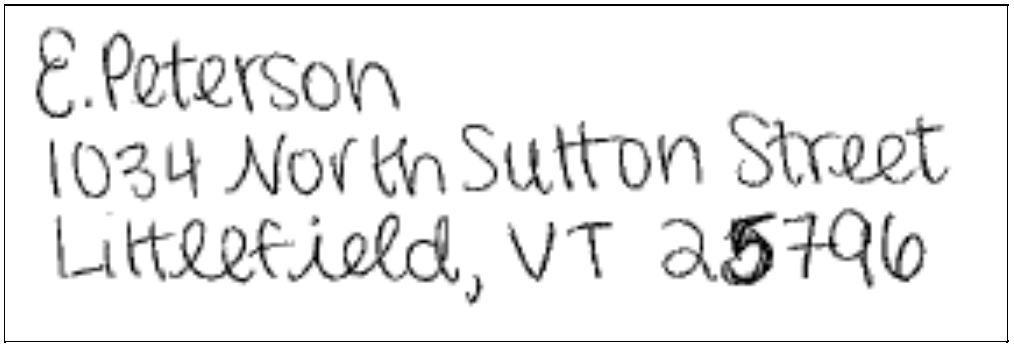
Can you reliably tell whether handwriting belongs to a male or a female?



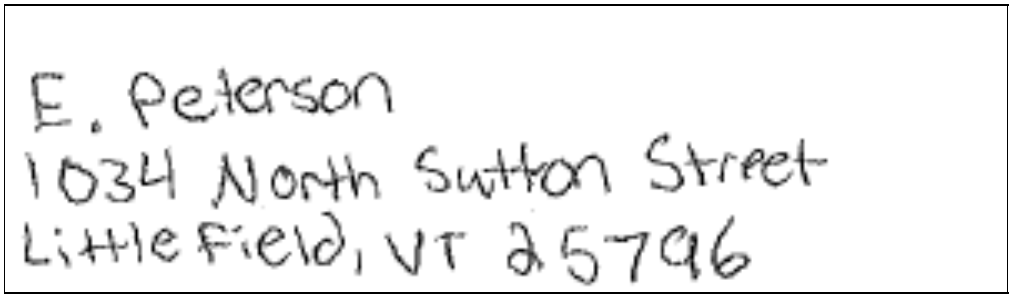
Handwriting samples

1. You will be shown 25 handwriting samples
2. For each, guess whether writer is
    - Male: write `0`
    - Female: write `1`
3. Write guess next to appropriate line on paper.
4. I'll give you ~5 seconds per sample.

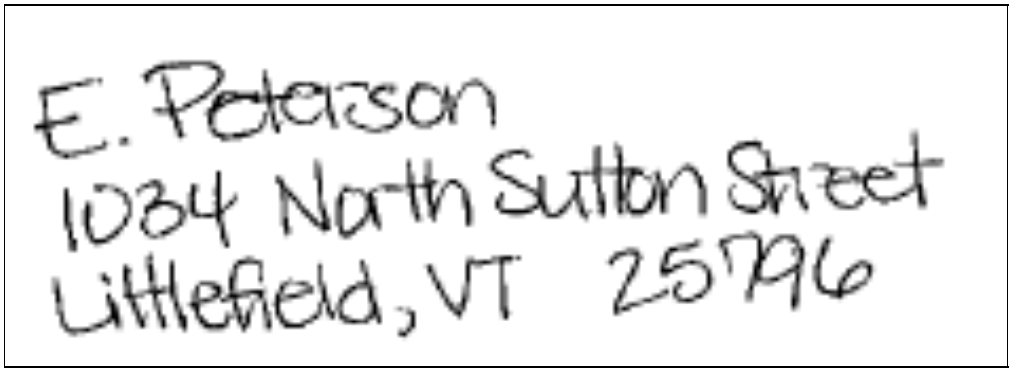
Sample 1



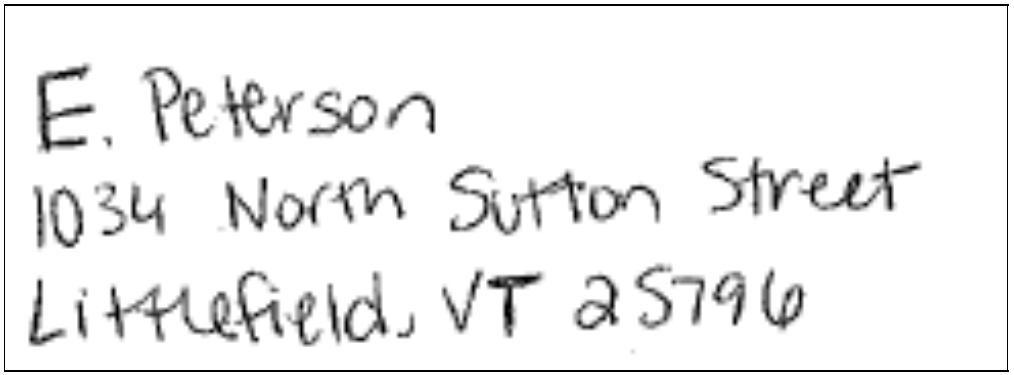
Sample 19



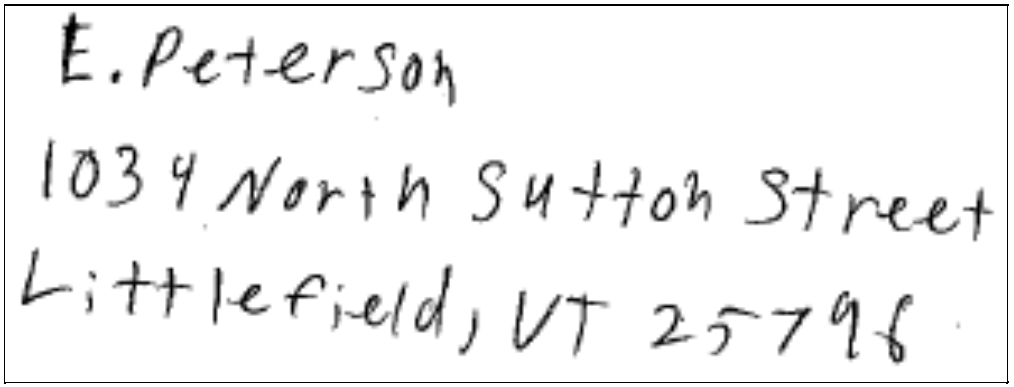
Sample 4



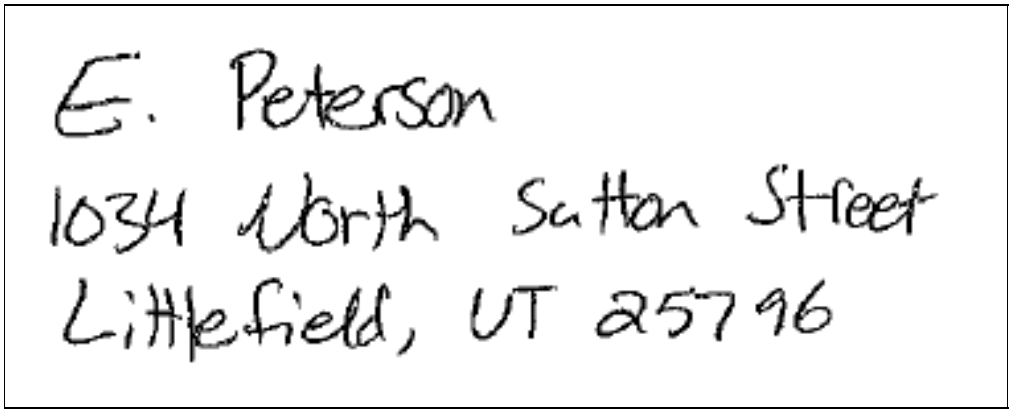
Sample 18



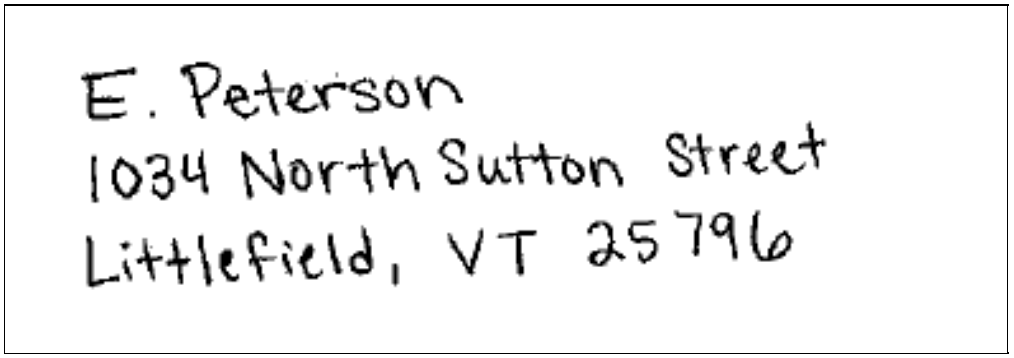
Sample 14



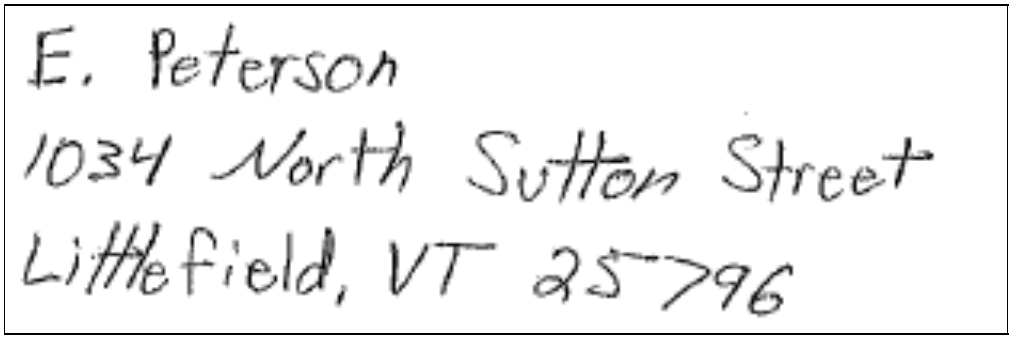
Sample 21



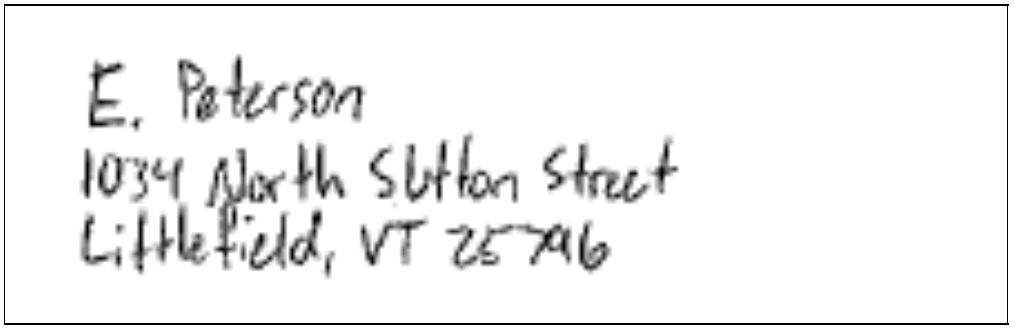
Sample 24



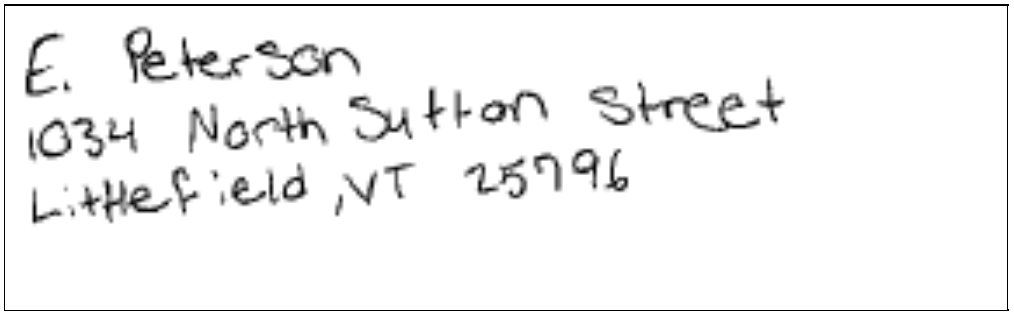
Sample 17



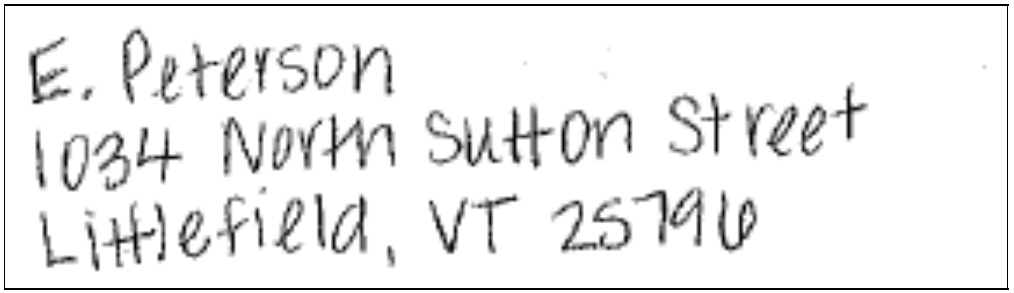
Sample 8



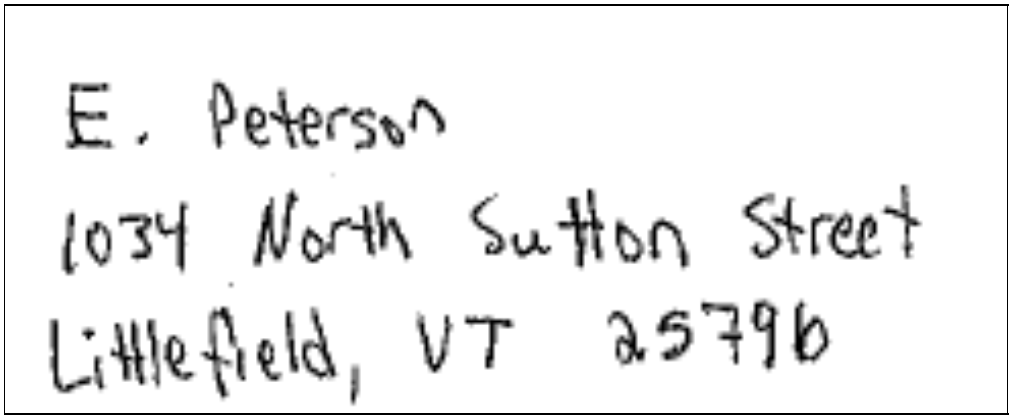
Sample 20



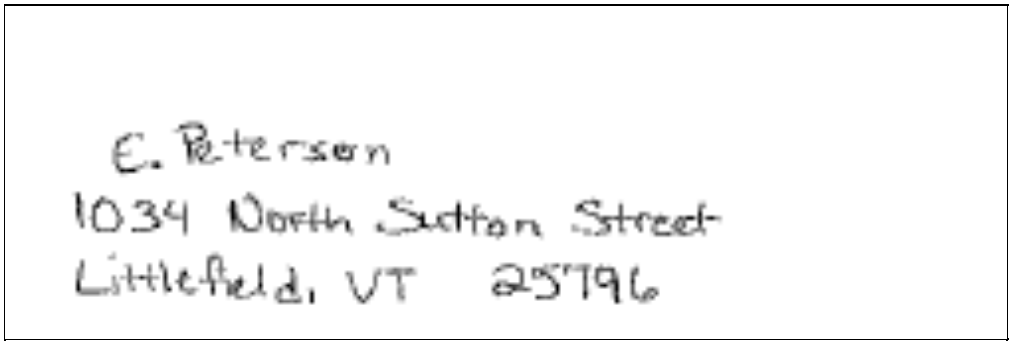
Sample 2



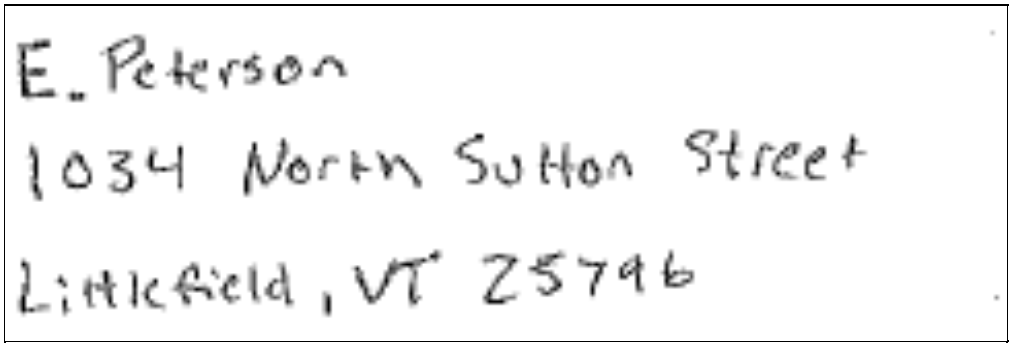
Sample 7



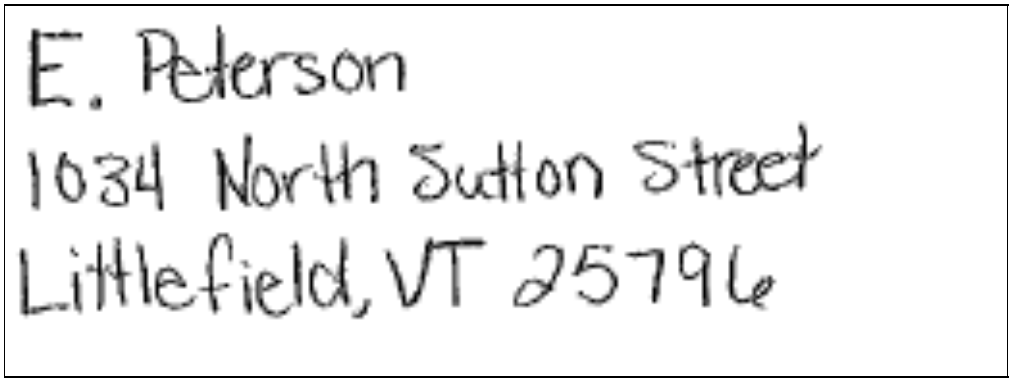
Sample 10



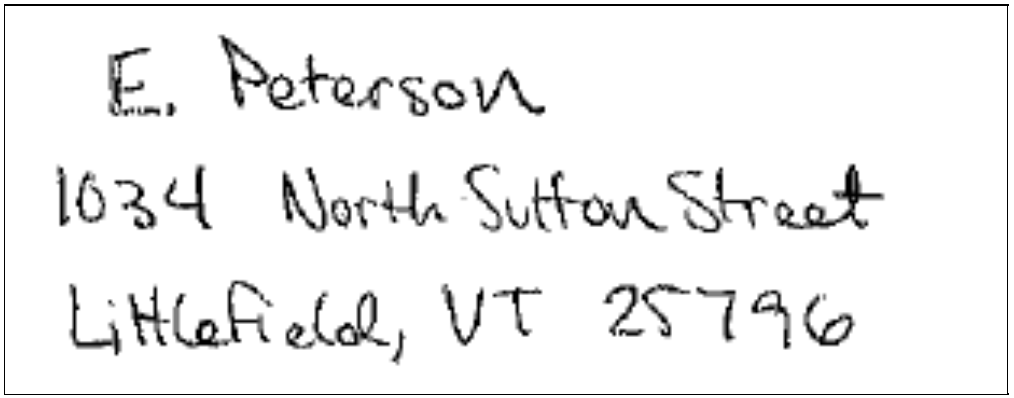
Sample 16



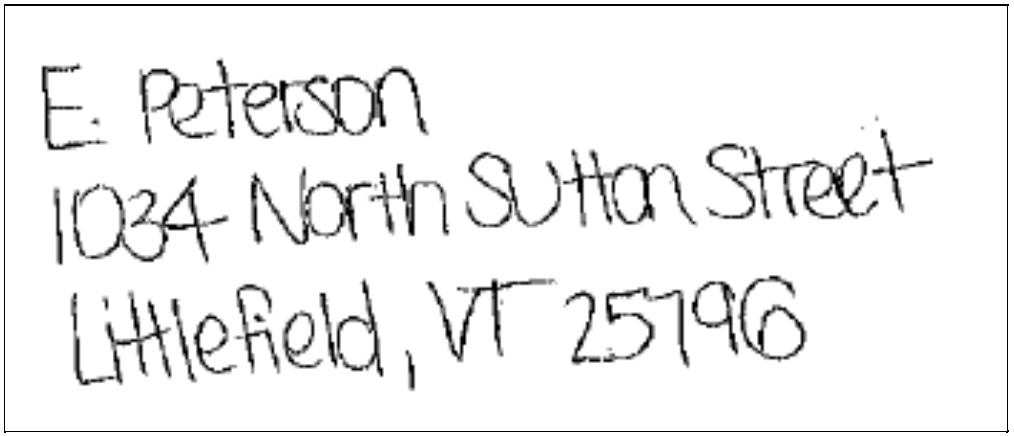
Sample 9



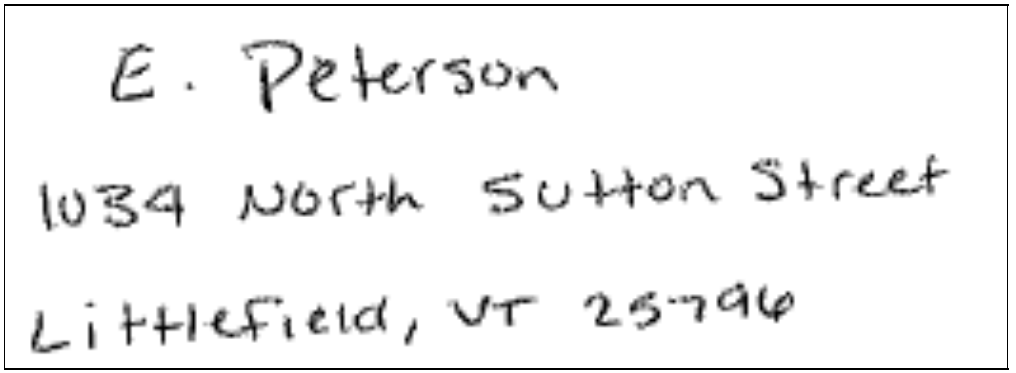
Sample 22



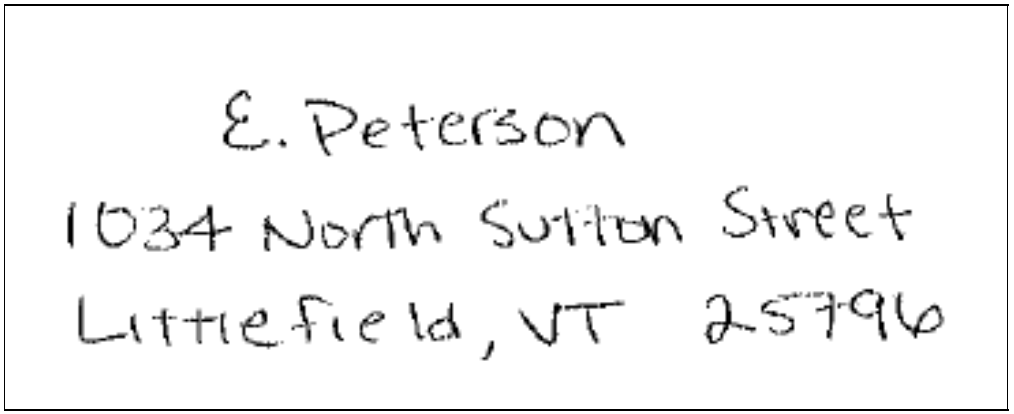
Sample 25



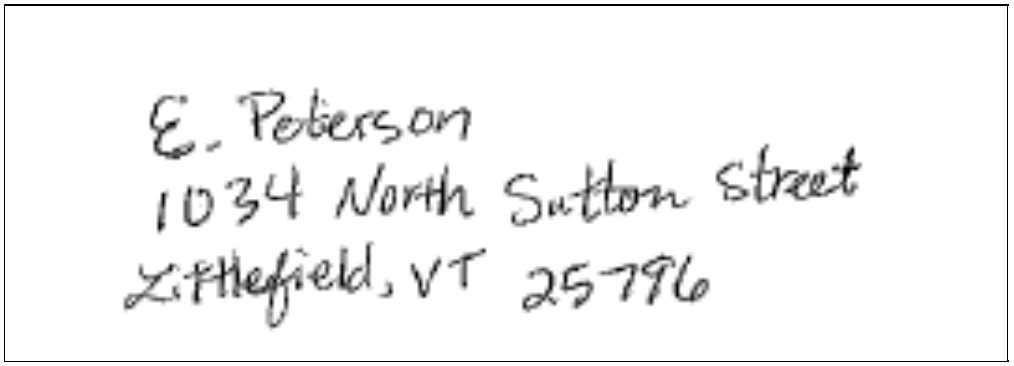
Sample 5



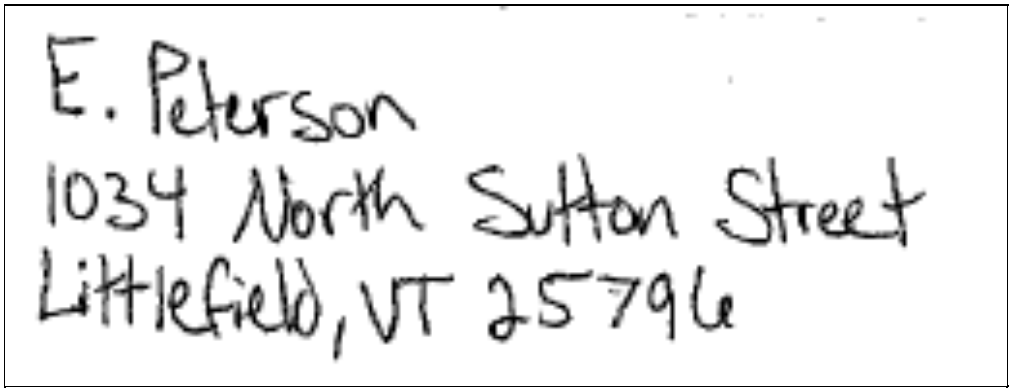
Sample 23



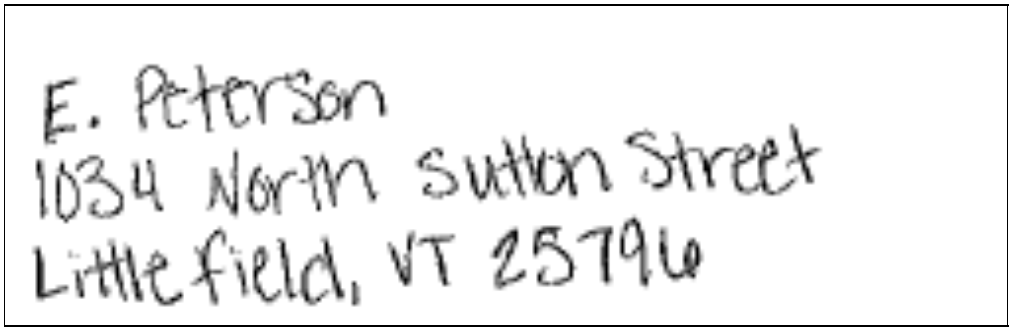
Sample 6



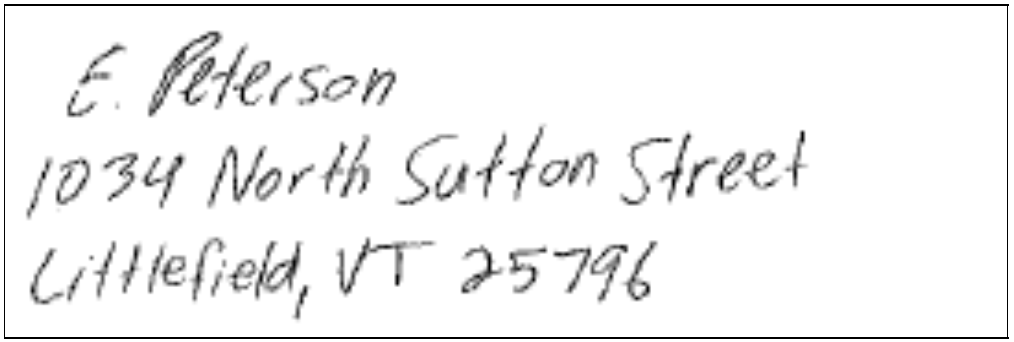
Sample 12



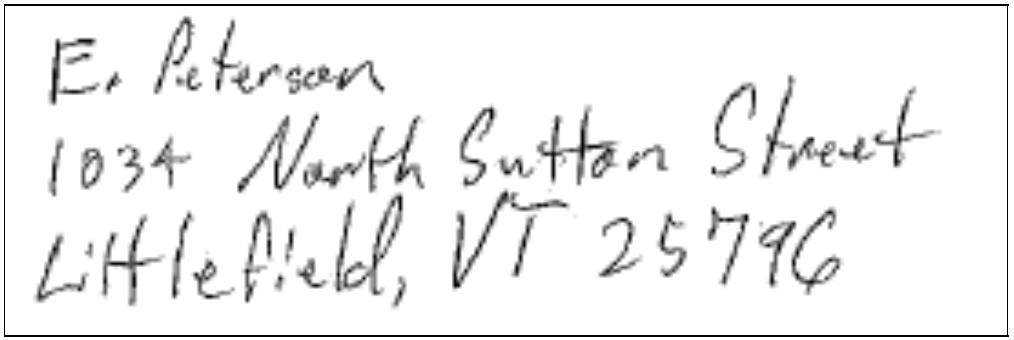
Sample 13



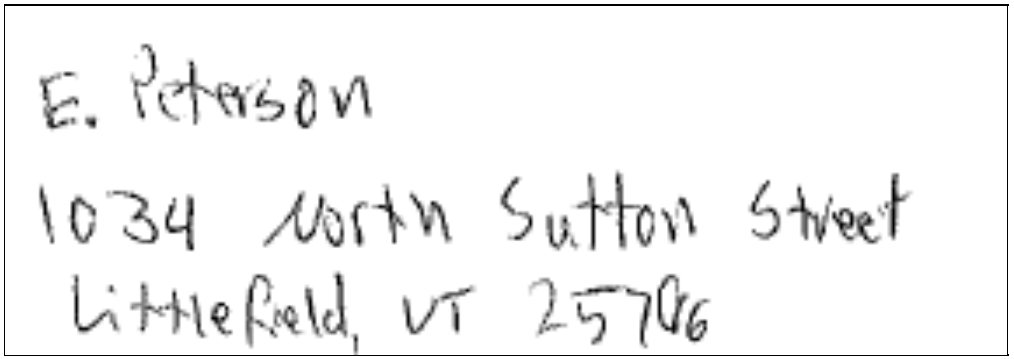
Sample 11



Sample 3



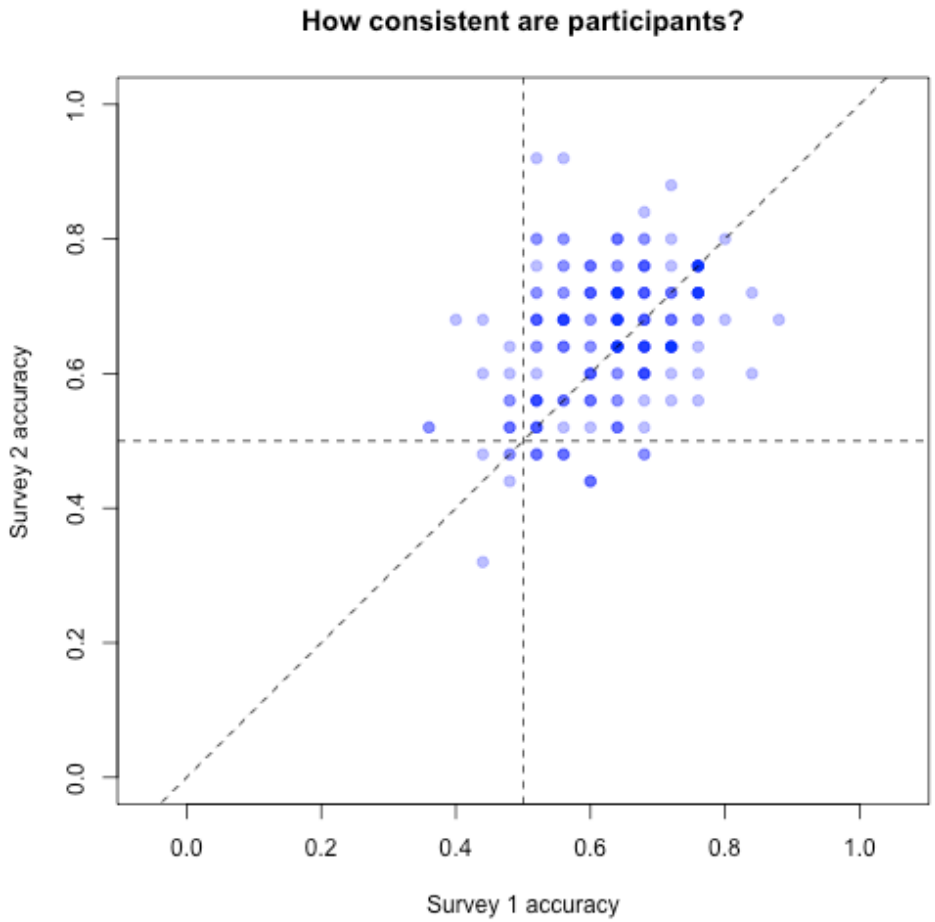
Sample 15



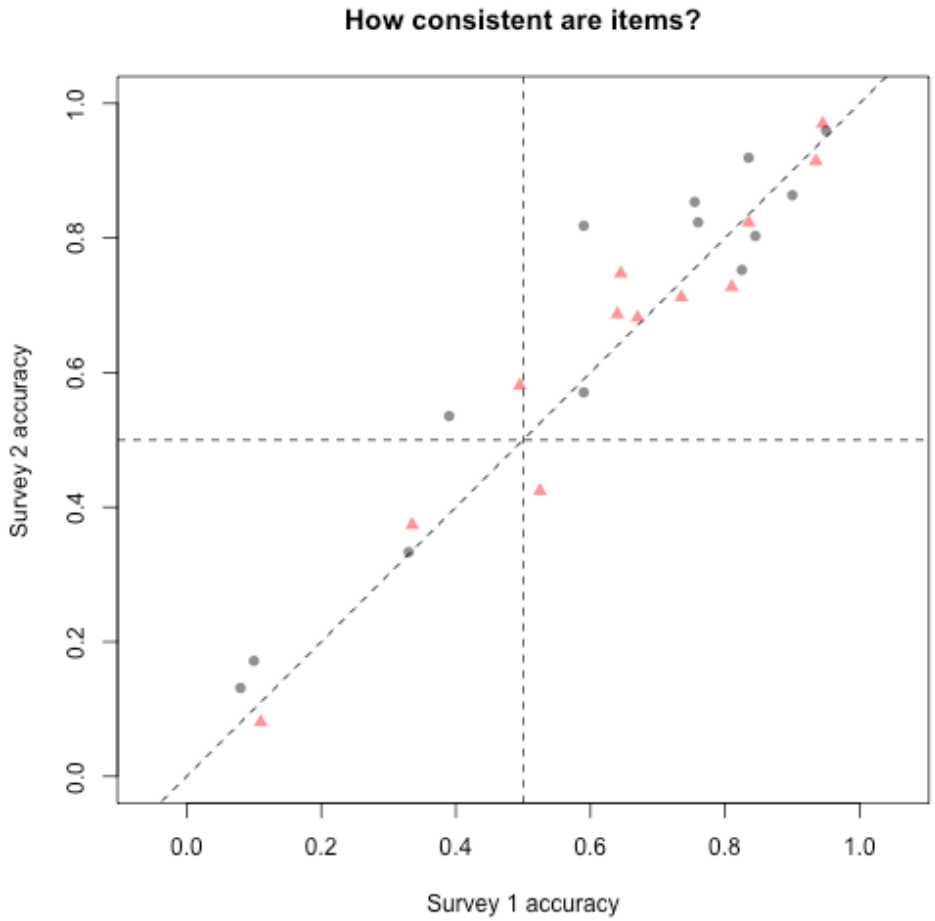
Predictions?

- If people can distinguish them, individual participants' tests should correlate.
- If there is a signal to distinguish, it should vary across items consistently.
    - A writing sample high in femininity/masculinity should provoke consistent responses.

Results


How consistent are participants?
Survey 2 accuracy
Survey 1 accuracy
0.0
0.2
0.4
0.6
0.8
1.0

Results

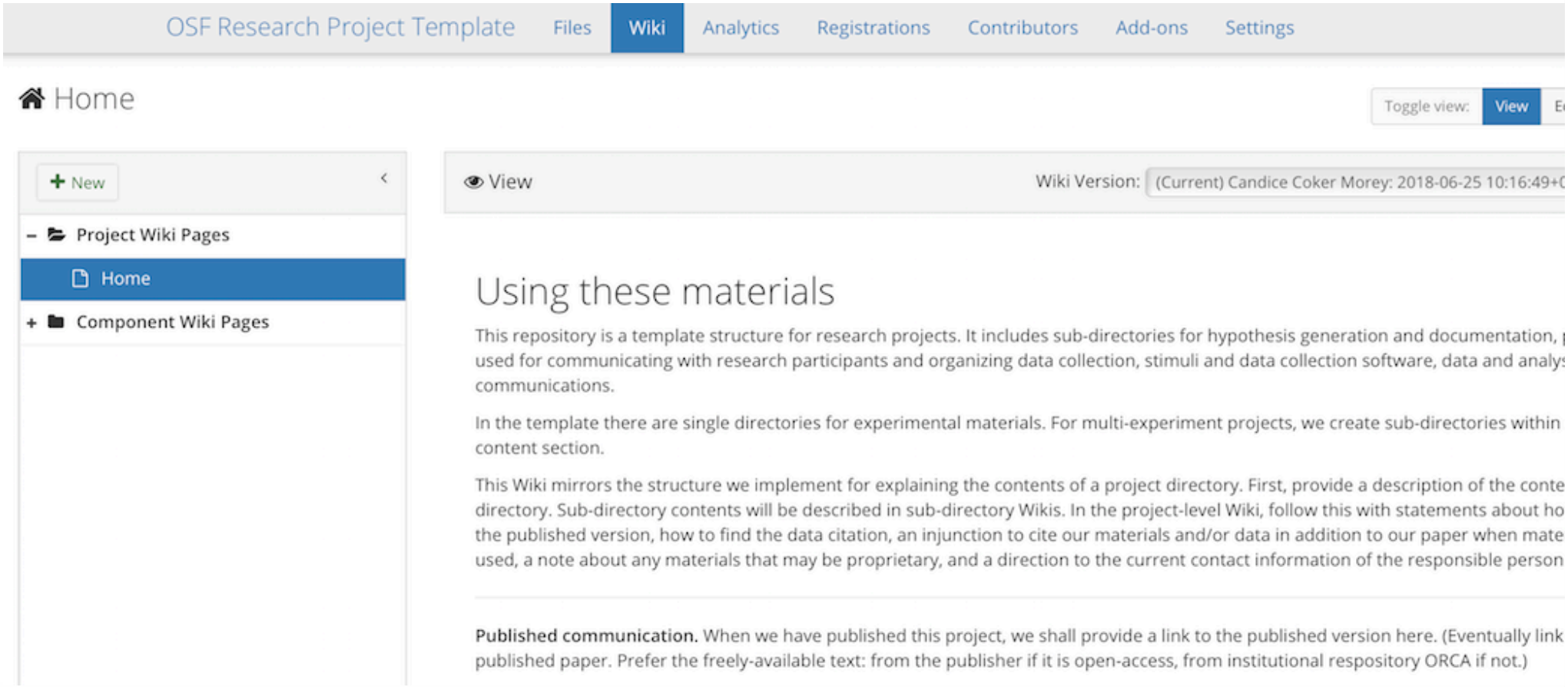

How consistent are items?
Survey 2 accuracy
Survey 1 accuracy
0.0
0.2
0.4
0.6
0.8
1.0

Creating our OSF project

- You've carried out a study about handwriting identification.
- Now organize your materials in an OSF project.
- Feel free to work in teams or on your own.
- Keep it simple!
- I'll show you an example.
- Highlights: directory structure, Wikis, contributors, templates vs. forking, guidance

Our project Wiki

- The first thing a visitor sees! https://osf.io/h6ekp/



Working on OSF pages

- Add me as a contributor if you want me to see how you did it, and maybe demo your work before lunch
- Next step: We will be thinking of a simple extension to be pre-registered. Think about what you might change to generalize or extend this study.

Registration: A frozen, time-stamped version of your project

- Open-ended registration: All-purpose registration tool
  - Prompts you for a short text describing what is being registered
  - E.g., "This registration records the state of the project before data collection began."
- Try it! Go to the Registrations tab . . .
- Now no matter what changes about your project later, you will know its state at this time point.

When to register?

- Can think of it like a back-up
- Use it to record the project's state at major milestones. Like
  - Right before data collection starts
  - When you have finished analyses or submitted paper

New registrations

Registrations

There have been no completed registrations of this project. You can start a new registration by clicking the “New registration” button, and you have the option of saving as a draft registration before submission.

New registration

For a list of the most viewed and most recent public registrations on the Open Science Framework, click here.

Kinds of registration?

Register

Registration creates a frozen version of the project that can never be edited or deleted but can be withdrawn. Your original project remains editable but will now have the registration linked to it. Things to know about registration:

- Ensure your project is in the state you wish to freeze before registering.
- Consider turning links into forks.
- Registrations can have embargo periods for up to four years. If you choose an embargo period, the registration will automatically become public when the embargo expires.
- Withdrawing a registration removes the contents of the registrations but will leave behind a log showing when the registration was created and withdrawn.

Continue your registration by selecting a registration form:

- ○ Prereg Challenge
- ◉ Open-Ended Registration
- ○ AsPredicted Preregistration
- ○ Registered Report Protocol Preregistration
- ○ OSF-Standard Pre-Data Collection Registration
- ○ Replication Recipe (Brandt et al., 2013): Pre-Registration
- ○ Replication Recipe (Brandt et al., 2013): Post-Completion
- ○ Pre-Registration in Social Psychology (van 't Veer & Giner-Sorolla, 2016): Pre-Registration

Cancel | Create draft

When to click the "Make Public" button

- Depends!
- Do your data contain "personal" information?
  - Then maybe never
  - Can still make rest of project discoverable
  - Can encrypt it
  - Try to avoid this barrier going forward (e.g., with explicit consent when appropriate)
- Making data anonymous
  - No birth dates, addresses (land or web), names, meaningful ids
  - Maybe - no dates, times

Advice on making data public

- My worst experience sharing data was actually a good one.
- It is convenient to have your data in public.
  - You cannot lose them.
  - Colleagues can get them without bugging you.
- Colleagues wanting to see your data means they are interested.
- What about "scooping"?

Planning an extension

- Brain-storm: Suppose we want to continue this project. Suppose we're skeptical of the results. How could we test whether they generalize?

Lunch provided by

After lunch: pre-registering a new project using OSF

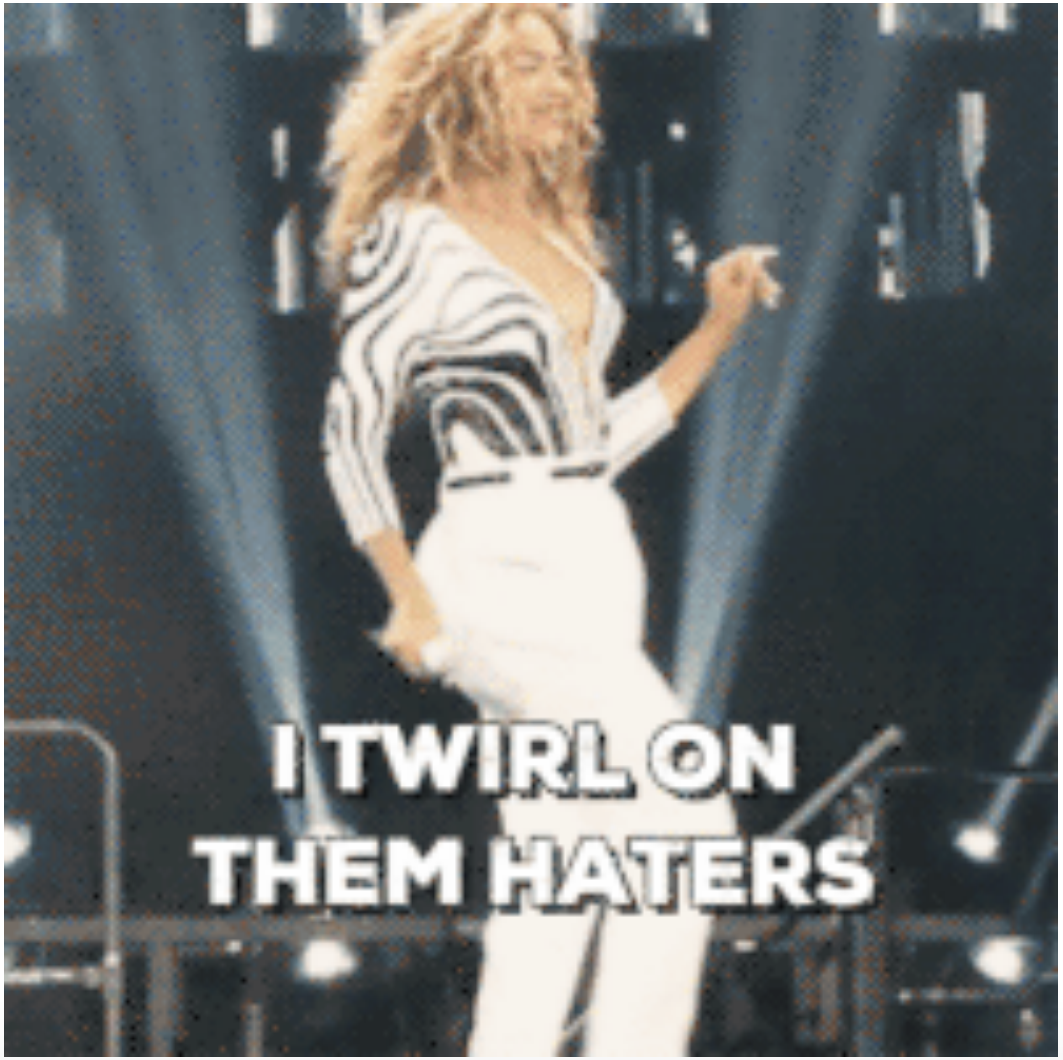
III. Transparency from the beginning: Pre-registering

The problems pre-registration solves

- Knowing what results mean depends on what you believed when you set up the study.
- Confirmation is far more convincing than exploration so we must know which tests were confirmatory.

My first reaction to these problems

My first reaction . . .



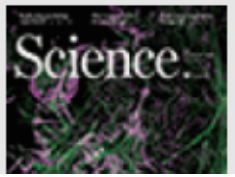
. . . was wrong

Basic research doesn't always replicate

RESEARCH ARTICLE

Estimating the reproducibility of psychological science

Open Science Collaboration*,†



Science
Vol 349, Issue 6251
28 August 2015
Table of Contents

Basic research doesn't always replicate

	Replications $P < 0.05$ in original direction	Percent
Overall	35/97	36
JPSP, social	7/31	23
JEP:LMC, cognitive	13/27	48
PSCI, social	7/24	29
PSCI, cognitive	8/15	53

- Could we be HARKing? (Yes)
- Could we be p-hacking? (Yes)
- Could we be presenting exploratory analyses as confirmatory? (Probably)
- Most of us have done some of these (John, Loewenstein, & Prelac, 2012)

Using pre-registration

- Pre-registration: Specify beforehand 1) method, 2) analysis plan, 3) what you will look for in your data to make your argument
- Tool to prevent yourself from HARKing and incidental p-hacking. Ideas:
  - Specify stopping criterion
  - Limit outlier removal
  - Map DVs to interpretations
- There is not only one way to use pre-registration!

Our next study?

1. Brainstorm.
2. Craft a simple text explaining our new study and expectations.
3. Add it to our project? Fork our project into a new project? Let's discuss.
4. Register.

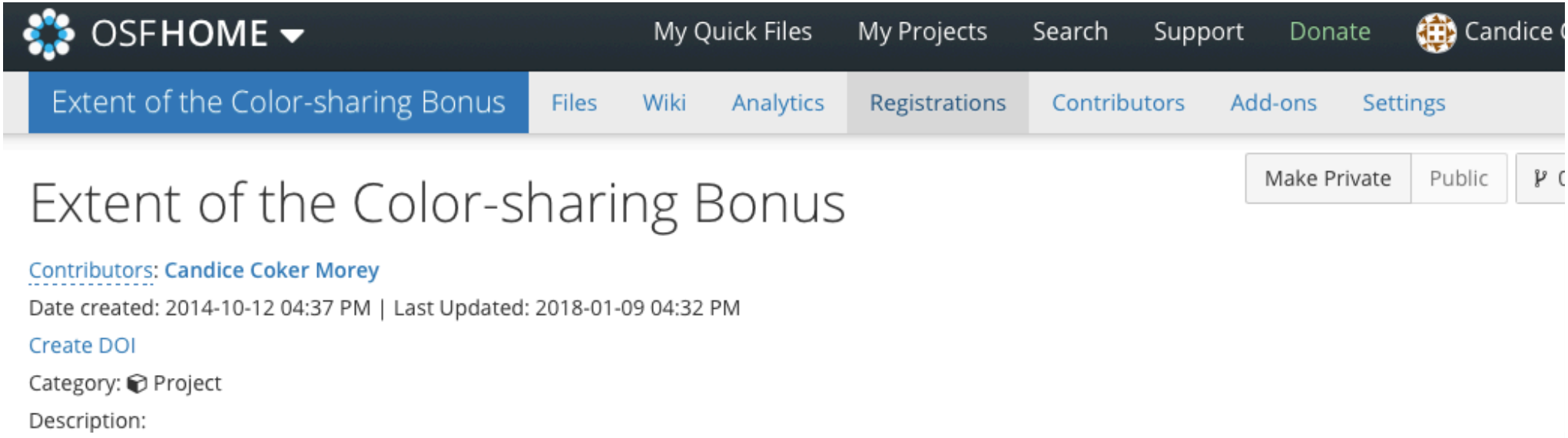
The spectrum of pre-registration options

- Low- or high-tech
- Quite informal to a contract with an editor
- Involve minimally changing the order of one step, possibly flipping the order of more steps

The most rigorous, formal choices might suit some projects but not others.

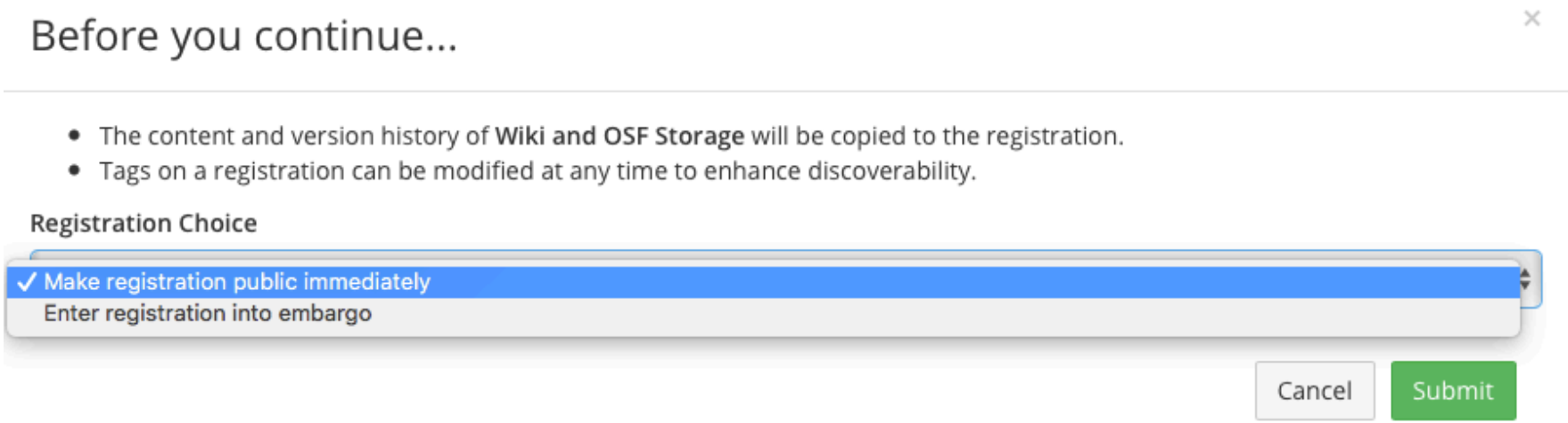
But documenting your pre-data hypotheses is always a good idea.

How "registration" differs from the project



- Creates a frozen version of your project's state. Can't be changed, though project can.
- The frozen version is dated.

Registration not necessarily public

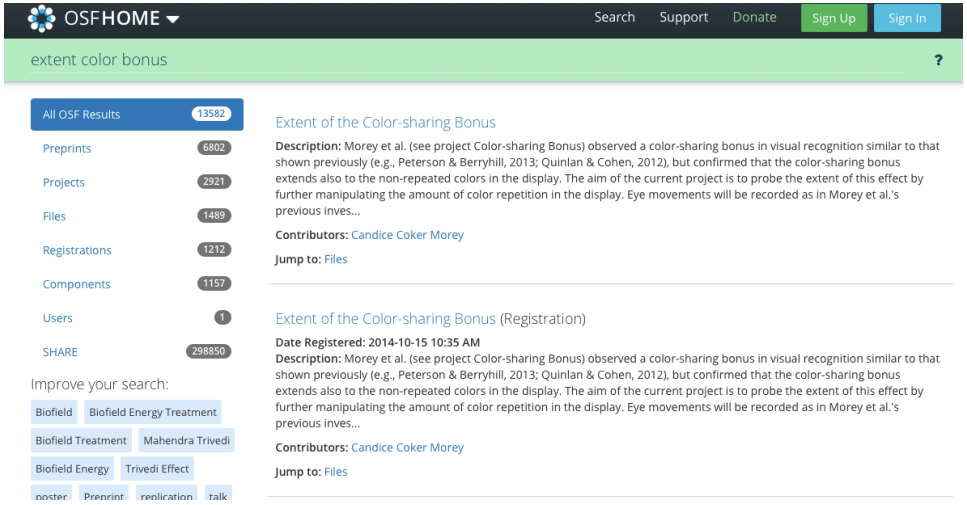


- Registrations become public after embargo (can be up to 4 years), or when you switch it to "public".
- Not a privacy issue if your registered project page excludes sensitive data.

Publishing a pre-registered project

- Make registration public
- Provide a link in your paper
- Readers will be able to examine the elements of the project that were registered

Example: Current versus frozen project



- Shows reviewers: Elements frozen on 2014-10-15 include hypothesis, analysis plan, the to-be-run experimental software
- Is it "pre"? Additional elements on project (if public) will be dated later than the frozen registration.

Open-ended pre-registration is

- Easy - requires no additional planning, relies on point-and-click web-based tools
- Flexible - can be a single paragraph just to provide evidence that hypothesis pre-dates data, can be more detailed
- Quick - minimal steps include very little beyond what you normally do (e.g., write a paragraph about what you will do)
- Helpful - for communicating about plans and hypotheses within own research group
- But does it ultimately aid publication?

Pre-registration and that reviewer who can't be satisfied

The Registered Report

Project is considered by the journal in two phases

- Stage 1: You submit your Introduction, Methods, and Analysis Plan for peer review. Data have not been collected.
- There are no results yet. Reviewers must focus on the rationale and design of study, the appropriateness of analysis plan.
- Editor may ask you to revise your Method based on reviewers' suggestions.
- Stage 1 manuscript gets in-principle acceptance. This part is frozen.
- You collect the data, run the planned analyses, and finish the Stage 2 paper.
- Stage 2 paper can only be rejected if you did not follow the approved Stage 1 plan.

Could also call this a "flipped" paper.

Yes! Let's do that!

- You only have to persuade about the rationale and design.
- Allows you to plan a risky project: Results are not known during evaluation.
- Reviewers: Much easier to review a prospective project.
- Team of researchers: Reward is known before bothering with data collection.

No way!

- Too much work!
- Interpretation of results could be embarrassing.
- Takes too long to start project.
- What if I find something that I didn't predict?
- What if I could have published in a better journal?
- Editors: What happens when results don't make sense?

There's something interesting that I failed to predict!

- You can still report it! That's no problem.
- You cannot re-frame your introduction to make it seem like you predicted it all along. No problem, right?

Not more work - different sequence of same work

- You do this now:
  - Plan experiment
  - Collect data
  - Analyze data
  - Write manuscript
  - Persuade journal

Not more work - different sequence of same work

- Registered report:
  - Plan experiment
  - Write 75% of manuscript
  - Persuade journal
  - Collect data
  - Analyze data
  - Write remainder of manuscript

Does it take more time?

- Be sure to consider the same end points!
- End of RR procedure: Paper is published
- End of usual approach: Data are (partly) analyzed, manuscript is (partly) written, disposition remains unknown

What if results do not make sense?

- If you want to do an exploratory project, writing an RR is not appropriate.
- If you want to do a confirmatory project, what's the problem?

As editor: RR proposals are clear

As author: Preparing an RR is hard, enlightening

- Currently preparing an RR and an RRR (Registered Replication Report), both with collaborators
- Did not realize how many details are usually decided in the moment, or by whoever is programming
- Perhaps that's more time spent planning?
- But: Less time fixing errors caused by running confounded or less-than-optimal experiments
- The constraint is intimidating.
- I will not miss the search for a publisher.

Hesitations about pre-registration?

- What about projects on a fixed time frame?
- What about time- or resource-limited data collection?
- What about student-led projects? What if our predictions differ?
- Not the only good way to persuade that your data say what you say they say (see also multi-verse analysis)

Summary: Working transparently

- Documenting your work in the open
  - Is increasingly expected
  - Can ease communication among and outside your group (and with Future You)
  - Can help you benchmark your progress
- OSF is a tool for documenting and communicating the work behind your paper.
  - This can be totally public or selectively shared
  - Registrations can document project state at key points
  - Registrations can legitimize confirmatory statements
- Pre-registration, even if basic, is a tool to keep yourself honest about what was confirmatory.
  - Restricts ability to use post-hoc re-framing
  - Will make our work more robust and cumulative

Helpful resources for getting even better

- Guide to transparency (Klein, et al., 2018)
- Try RStudio and markdown for analysis

Thanks for your attention!

Editor-in-chief, www.journalofcognition.org (We consider registered reports and we like pre-registration!)

My data and materials are publicly available on Open Science Framework (https://osf.io/4xwa8)

Blogging at The Mnemonic Lode, candicemorey.org

Twitter: @CandiceMorey