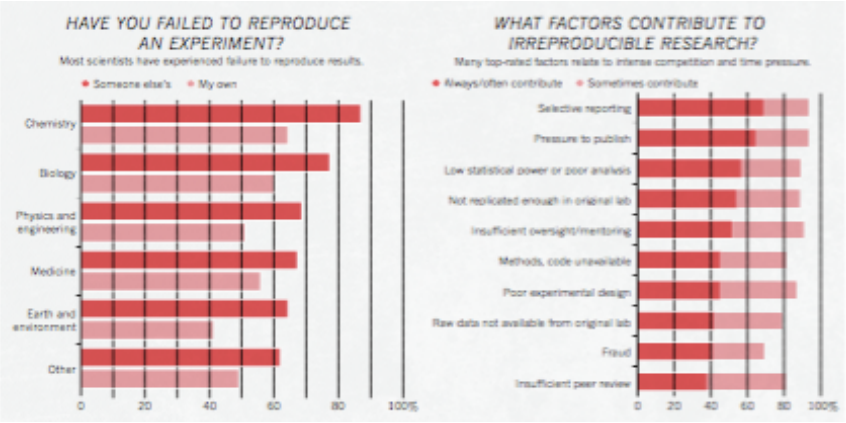
Research methods 13

Replicability in Psychological Research & Choices in Research

Caryn Block
ORLJ 5040
Teachers College
Columbia University

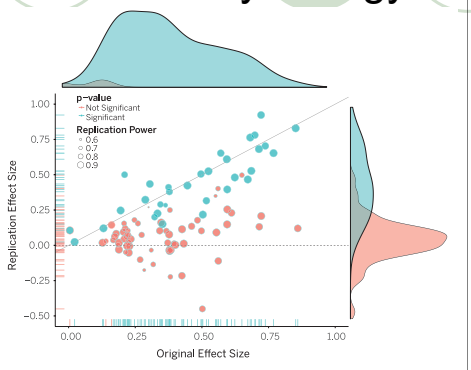
1

Replication crisis in science



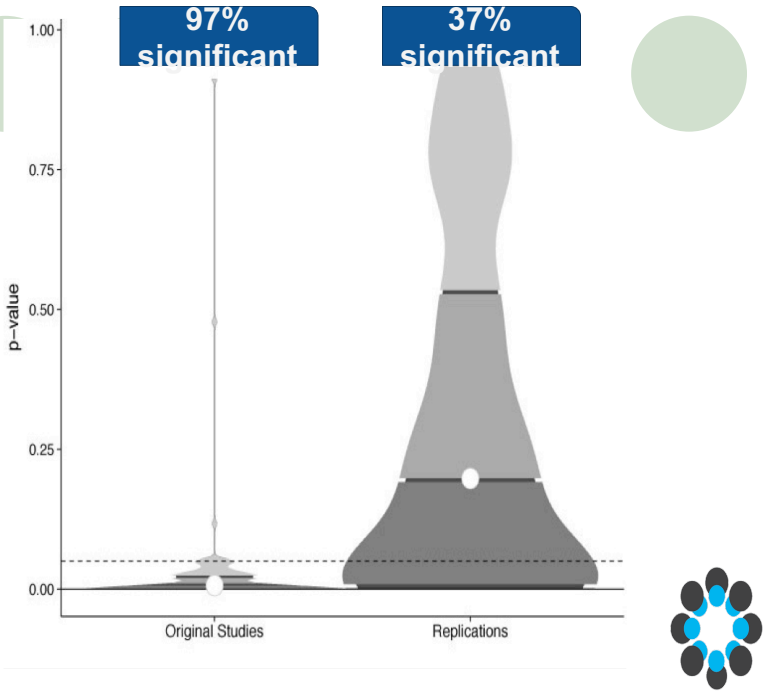
2

Open Science Collaboration (2015)
Replication Crisis in Psychology



Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.

3



4

By discipline within psychology

Journal	% Findings Replicated
Journal of Personality and Social Psychology: Social	23
Journal of Experimental Psychology: Learning, Memory, and Cognition	48
Psychological Science, social articles	29
Psychological Science, cognitive articles	53
Overall	36

5

Are these findings overstated?

- Gilbert et al (2016)
  - **Sampling issues in studies**
    - Replicators chose **which studies** they would replicate
    - Replicators **only used the last study** in a multi-study article
  - **Fidelity issues in studies**
    - Replicators used **different samples** that didn't always make sense (e.g. measuring attitudes toward African-Americans in an Italian sample)
    - Replicators used **low fidelity protocols** (e.g. American sample asked to anticipate consequences of honeymoon, Israeli sample asked to imagine consequences of military service)
    - Replication studies where **original authors endorsed the replication procedures** were 4x more likely to replicate (59.7% vs. 15.4%)
  - **Statistical issues in studies**
    - How many replications would **fail by chance** alone?
    - Replication studies were **under-powered** (see Maxwell et all, 2015)

6

Are these findings overstated?

- Van Bavel et al. (2016)
  - Ignored **contextual differences ("hidden moderators")** between original study and replication study
  - Found that even when they controlled for significant predictors of replication success found in the original studies, **contextual sensitivity predicted replications success**
  - Calls into question issues of **sampling** when conducting a "direct replication"?
  - Importance of creating replications that are high on **experimental realism rather than mundane realism**

7

Types of replication

- **Direct/Exact**
  - Recreate *physical* conditions used in previous research
- **Conceptual**
  - Recreate *theoretical* conditions used in previous research

8

Even if overstated, they've cast a negative cloud over psychology

9

Why are we here?

- **Increased visibility in popular press**
- **Pressure to publish**
  - File drawer problem
  - Significant findings published more than non-significant findings
- **Lack of transparency about methods**
  - Researcher degrees of freedom
    - Collect more data?
    - Primary DV?
    - Excluding observations?
- **Misunderstanding of generalizability**
  - Research doesn't generalize from one study to another because so much unique – role of theory in cumulating knowledge
- **Lack of statistical power in single replication studies (Maxwell et al, 2015)**
  - Utility of meta-analyses to look at robustness of findings

10

Cohen (1990)
Things I Have Learned (So Far)

- **Less is more**
  - Except for N
  - Variables in a study
  - Reporting numerical results
- **Simple is better**
  - Reporting of data – graphing, scatterplots
  - Unit weighting

11

Cohen (1990)
Is p<.05 really a magic number?

- **Power analysis** – the cost of overvaluing Type I error
  - **alpha** (Type I error)
    - Probability of falsely rejecting null
    - Concluding there is an effect when there isn't
    - Typically held constant at .05
  - **N** (sample size)
  - **ES** (effect size)
    - Low, moderate or strong effect in the literature?
    - Can be correlation or d= X_1-X_2/sd
  - **beta** (Type II error)
    - Probability of falsely accepting null
    - Concluding there is no effect when there is one
    - Typically very high in most studies (.5-.8)
- Recognize both Type 1 & Type II error in your own research and make an informed decision

12

Solutions?

- **Plan your research**
  - Power analyses
  - Effect size measures are overestimated in the literature

    https://designingexperiments.com/shiny-r-web-apps/
- **Joint planning of replications with original researchers**
- **Pre-register your sample, methods, analyses plan**
- **Remember what can be learned in one study**

13

Latham, Erez, and Locke
Methodological Case Study

Common phenomena: researchers studying same area come up with opposite conclusions

- Usual method of resolving this: dog fight in journals or at conferences
- This study is an example of a better method of solving this problem
- Joint design of critical studies by antagonists
  - Look at operationalizations, instructions, and setting to figure out critical issues to design studies around
  - Design and run critical studies
  - Need a 3rd party as a mediator
- Great example of research process building knowledge
- Two antagonists still draw different conclusions!

14

Publishing Replications

- **Psychfiledrawer.org**: Archives attempted replications of specific studies and whether replication was achieved. http://psychfiledrawer.org/
- **Center for Open Science:** Open Science Framework, where replications can be reported. https://osf.io/
- **Association of Psychological Science:** Has registered replications of studies, with the overall results published in *Perspectives on Psychological Science*
- **Plos One: Public Library of Science:** Publishes a broad range of articles, including failed replications, and there are occasional summaries of replication attempts in specific areas.

15

Preregistration

[Ambassador Name]

[Ambassador Title/Affiliation]

Ambassador, **Center for Open Science**

cos.io/prereg

16

Questions for Discussion

1. Do the failures of replication shake your faith in what you have learned about psychology? Why or why not?
2. In what types of research is replication more important? When is it less important?
3. If a finding fails to replicate in one study should the results be discounted? If a finding does replicate in one study are the results of the first study "true"?
4. How should we cumulate knowledge?
1. What are the benefits of the "Open Science" approach? What are the costs?
2. Why do you think this conversation has become toxic?
3. How should we move forward?

17

Wood & Wilson (2019)
No crisis but no time for complacency

National Academy of Sciences, Engineering and Medicine Committee Findings

Distinction between reproducibility and replicability

- **Reproducibility**
  - Obtaining consistent computational results using the same data, methods, code and analysis (using original data)
  - Should always be the case – need to make data transparently available
- **Replicability**
  - Obtaining consistent results across studies that were aimed at answering the same question, each of which obtained it's own data (using new data)
  - More nuanced – "The goal of science is not, and ought not to be, for all results to be replicable." (p.28)
  - There is a tension between replicability and discovery
  - Helpful source of non-replicability
    - Limits of current knowledge
    - Uncharacterized variabilities in the system
    - Leads to new insights
  - Unhelpful sources of non-replicability
    - Shortcomings in the design
    - Inappropriate statistical inference (p-hacking, cherry picking)

18

Why do we do research?

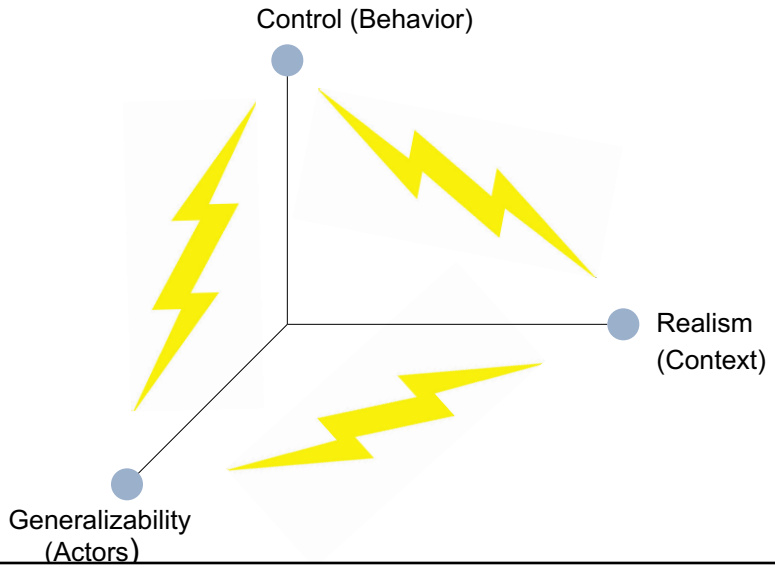
19

How is research different from other methods of knowing about human behavior?

- Less subjective
- Systematic
- Based on empirical process
- Subject to public inspection of findings
- Guided by ethical principles

20

The three **desiderata** of research



21

The Three Horned Dilemma

- Maximizing one or two desiderata will by definition minimize the others
- If you grab onto two of the horns, you are impaled by the third
- The research process is a set of problems, or choices
- All research is flawed

22

Which research strategy should you choose?

- Depends on your research question
- There is no ideal strategy—all have flaws

23

How to live with the **three-horned dilemma**

- The **research question** should determine the appropriate method for a given study
- A **theory is only as good as it's data** - so look for a balance between types of research that have been done to support it
- We "know" about human behavior when there are **theories** based on **multiple studies** using **multiple methods**
- Look for **convergence of findings** that grab onto different horns

24

Choices made in living with the **three-horned dilemma**

25

Rozin (2009)

What kind of empirical research should we publish, fund and reward?

- Balancing the desire for **rigor** in research with the desire for **relevance** in research
- Psychology is tipped too much toward rigor
- Need to be ask ourselves to what degree each study advances our work
  - An interesting study with methodological flaws may be more valuable than a flawless but uninteresting study

26

McGuire (1973)
The Yin and Yang of Progress in Social Psychology

- Inadequacy of theory-derived hypotheses tested experimentally in the lab
  - Create "**stage managers**" to confirm hypotheses
- Inadequacy of socially relevant hypotheses tested in the field
  - Create "**situation finders**" to confirm hypotheses
- Need to develop models and methods that mirror the multivariate complexity of cognitive and social systems
  - Allow ourselves to engage with "reality" rather than "shadows"

27

Kulka (1982)
Idiosyncrasy and Circumstance

- Rational process of doing research is **not nearly as rational** as we portray it to be, when writing research reports
- What else is driving our choices in research besides the three-horns?
- Idiosyncrasy and circumstance creep into **every stage** of the research process
  - Formulating the research problem
    - Resource availability
    - Political considerations
    - Becoming "expert" in a certain area
  - Designing the study
    - Research settings
    - Research strategies
  - Executing the study
    - Operationalizations of variables
    - Data collection process

28

Aronson (1995)
Research as a Leap of Faith

- An address given at a convention that is a response to the Wilson, Aronson & Carlsmith chapter
  - Empirical research described as methodical, slow, and perfectionistic
  - Yet, research always requires a leap of faith
  - More fun than they described
- **Bottom Line:**
  - Don't wait around for the perfect design
  - Take the leap and have fun
  - Ask the one question you really want the answer to!

29

Final thoughts about living with the **three-horned dilemma**

- Keep doing research!
- There is a lot more we need to know about human behavior
- You are the future of this enterprise
- Lead us into new directions and answer important questions in the best way that you can
- Remember that it is a gift to be able to study the issues that most interest you

30