The Integrity of Scientific Processes

**IF ACADEMIC INTEGRITY IS THE SOLUTION
WHAT IS THE PROBLEM?
SEMINAR 4 MARCH 2020**

Miles MacLeod
Dept. of Philosophy
The Graduate School

ITC
UNIVERSITY
OF TWENTE.

Academic Integrity Programme TGS (2019 +) for all PhD students

Plenary Introduction
(Peter-Paul Verbeek)

Session Academic Integrity
(Miles MacLeod & Patrick Smith)

Academic Integrity Online Course

(Miles MacLeod & Patrick Smith)

Final meeting

Goals Today:

Inform you about the goals and aims of academic integrity as we've developed it in this course -> how we are conceptualizing integrity.

But focus in particular on the integrity of science processes.

Why are we doing this course?

1. **Pressures on young academics** continue to increase, creating strong incentives for misconduct -> it's in your personal interest to have information and power (reasoning skills) to avoid these situations and make good decisions if they occur.

2. **Public scrutiny of science** itself is increasing following major fraud cases but also many dubious scientific claims in the media. Public trust in science is a serious issue in our context in which scientific claims are often contested -> and we all have a collective interest in the integrity of science.

What is the problem: Narrow Context

- A meta-analysis of survey studies asking researchers about scientific misconduct (Fanelli 2009) concluded that:
  - 2% admitted to having fabricated, falsified, or modified data or results at least once
  - 14% reported observing colleagues doing this
  - 33% admitted to other questionable research practices
  - 72% reported observing colleagues doing this
  - Because these are self-reports, the numbers are probably higher

- A study of journal retractions and corrections (Fanelli, Costas, and Larivière 2015) suggests that:
  - Misconduct is significantly higher in the earliest career stages
  - Men and women are equally likely to engage in misconduct
  - Misconduct is significantly higher in countries that lack research integrity policies and where mutual criticism is impeded

What is the problem: Broader Context

- Replicability crisis in social psychology and the medical fields.
- Concerns over the spread of contradictory research results in the public sphere.
- Concerns over the validity of big data and machine learning based methods given their rising influence and prominence.
- Worries about the replicability of complex simulation model results, and their validity.
- Worries about the underestimated methodological dependency of research results (particularly with respect to statistical methods).

Why is the *university* doing this?

Why us?

- Because universities are *responsible* for helping their students develop the understanding, ethical background, and skills and habits of analysis that will allow them to pursue research with integrity.

Why students?

- So students can be held responsible for your conduct and cannot in the future claim ignorance of integrity issues.

There's a code of conduct now…do we need this?

- An expansive (but strict) code of conduct now exists.

- Codes of conduct have their limits.

- We want students to be able make their own judgements about integrity even in unclear or borderline circumstances….and ultimately to protect both themselves and the scientific process.

Netherlands
Code of Conduct
for Research
Integrity

2018

Example 1

You've just finished your PHD thesis and in need of more publications in order to secure a position/grant. Your professor offers to take part of it (which isn't published yet) and write it up as full paper. He/she tells you that while this part is good and potentially publishable it requires work to get it into a publishable state and he or she has the experience to do this. It will take some work though so he/she would need to be first author. Would you accept this offer? Why or why not?

Example 2

A professor is running a project on the environmental causes of extremism and has collected a large amount of data tracking numerous economic, social and other behavioral variables, and political attitudes, in several European populations. As his PhD student he has asked you to go through each of those variables one by one, and find any links to political extremism which pass a significance test. You find some results. He suggests together you should publish immediately, proposing various potential causal relationships, and then publicise these results in the media.

Example 3

You are part of a project group examining the relations between water usage and the spatial distributions of populations in the third world in order to optimize water management. For your PhD you are doing a resource management modeling component of the project. The spatial distribution data you received is relatively smooth compared to that you have seen in other studies. You ask the other PhD students involved and each say that the professor produces these distributions not them. When you ask the professor how these results are produced she says it is complicated and as a modeler you don't need to worry too much where the data comes from. Indeed you should be happy the data is as smooth as it is.
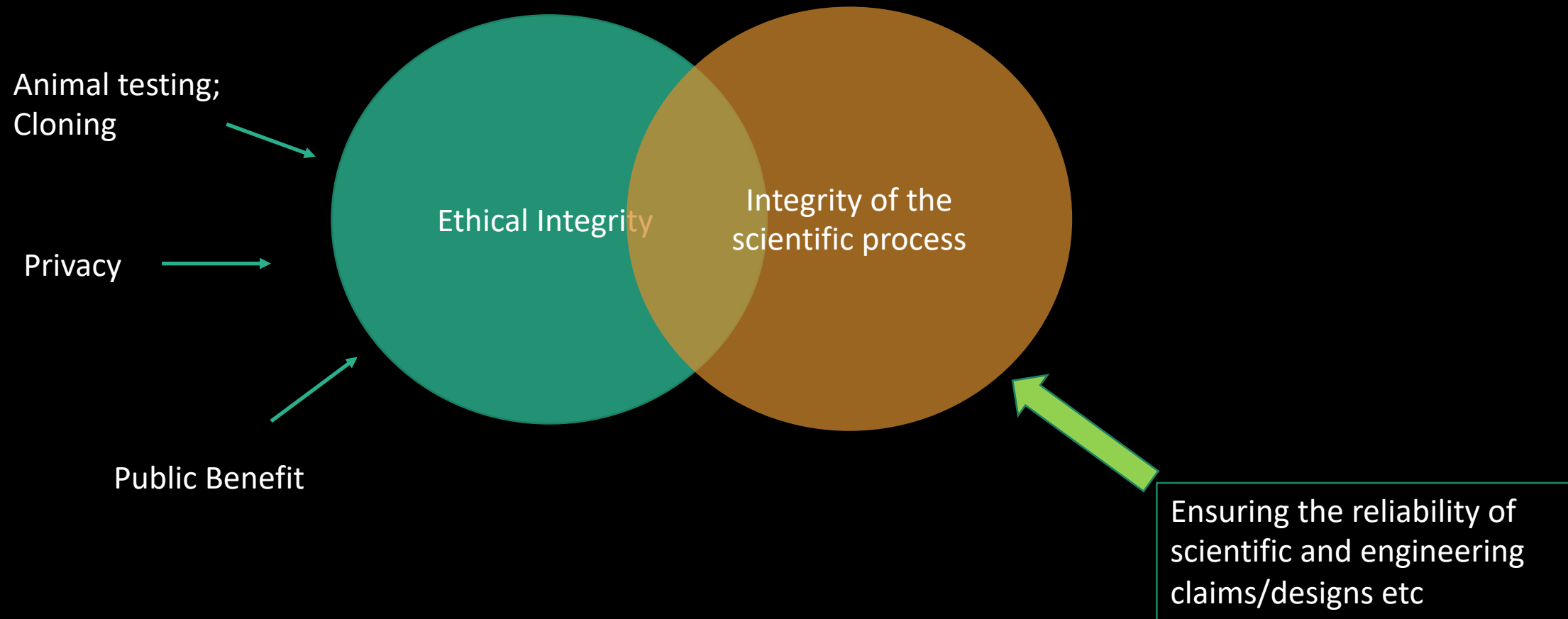
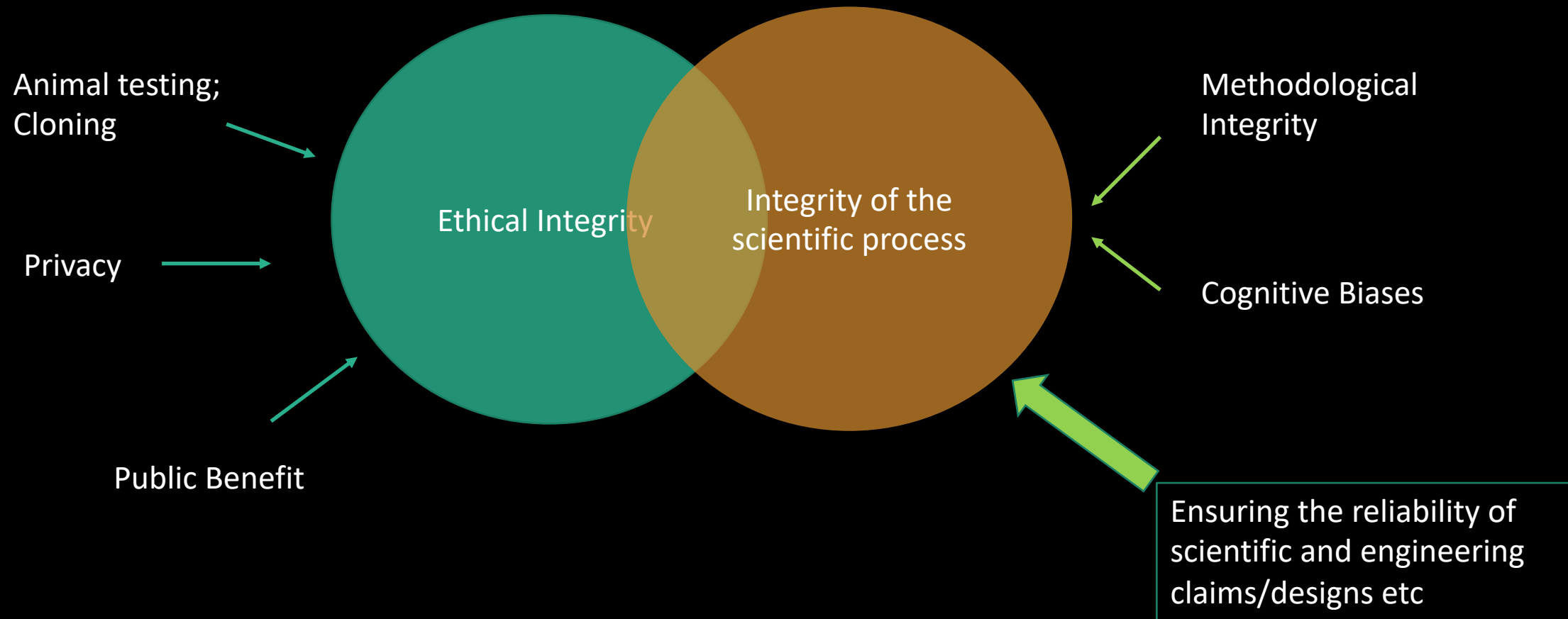# So how can we conceptualize integrity in the course?
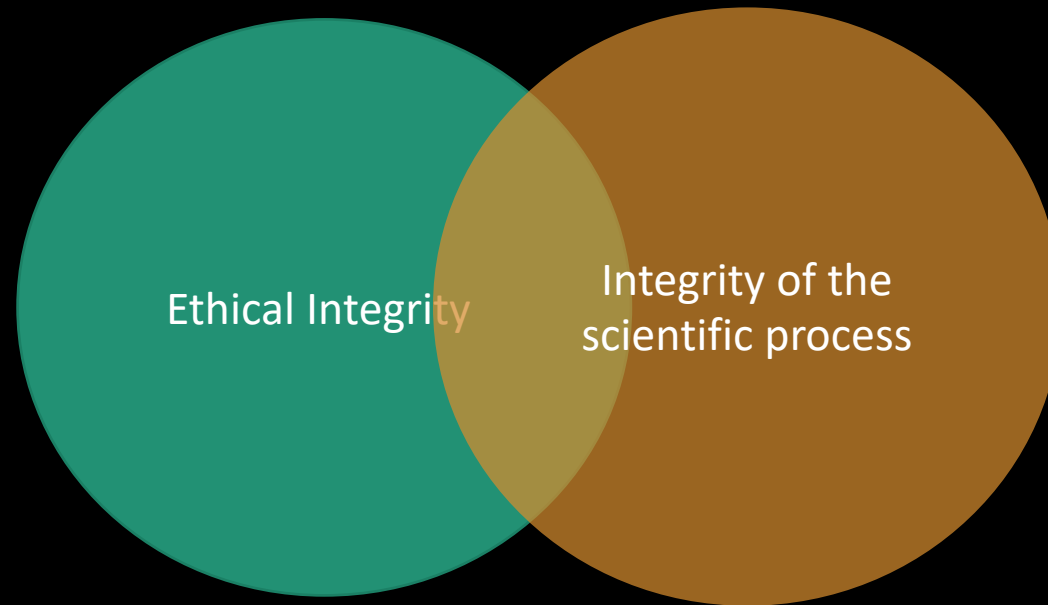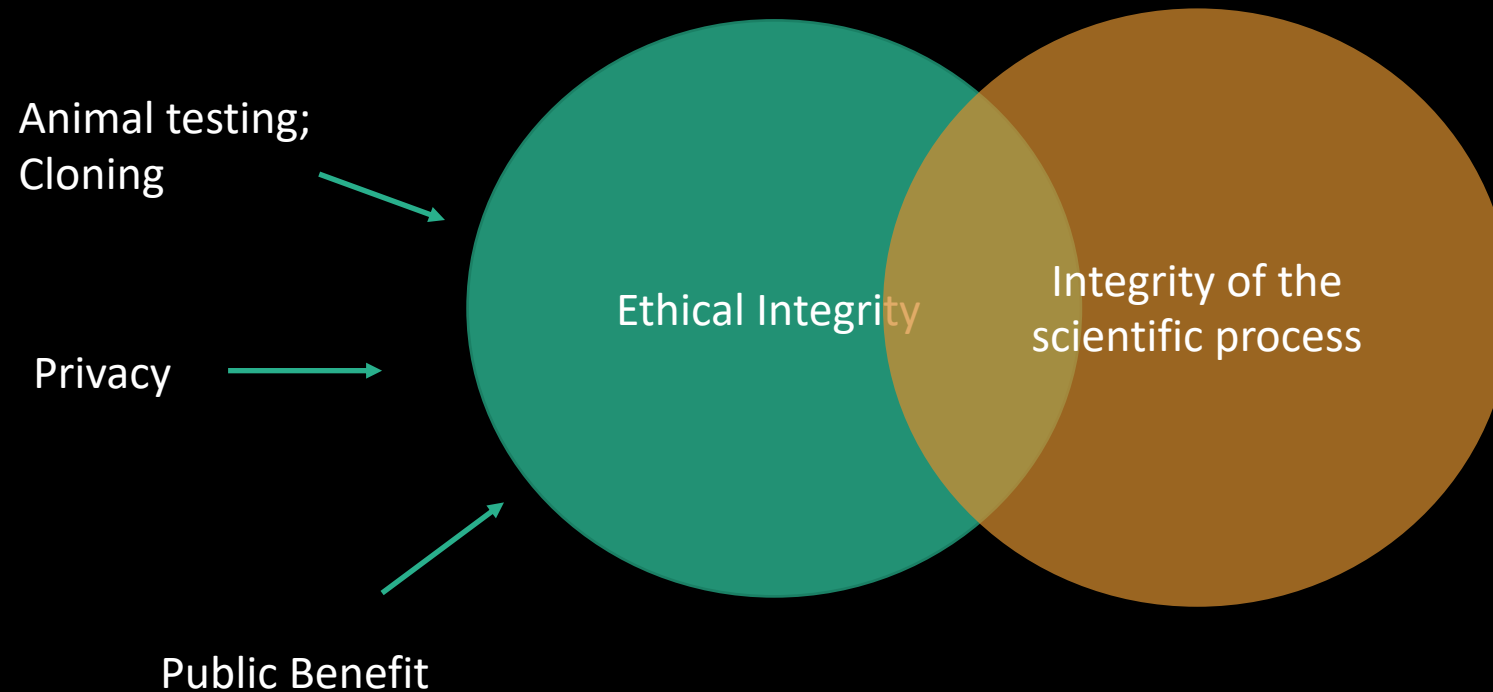
Ethical Integrity

Integrity of the scientific process

Components of Academic Integrity

Animal testing;
Cloning

Privacy

Public Benefit

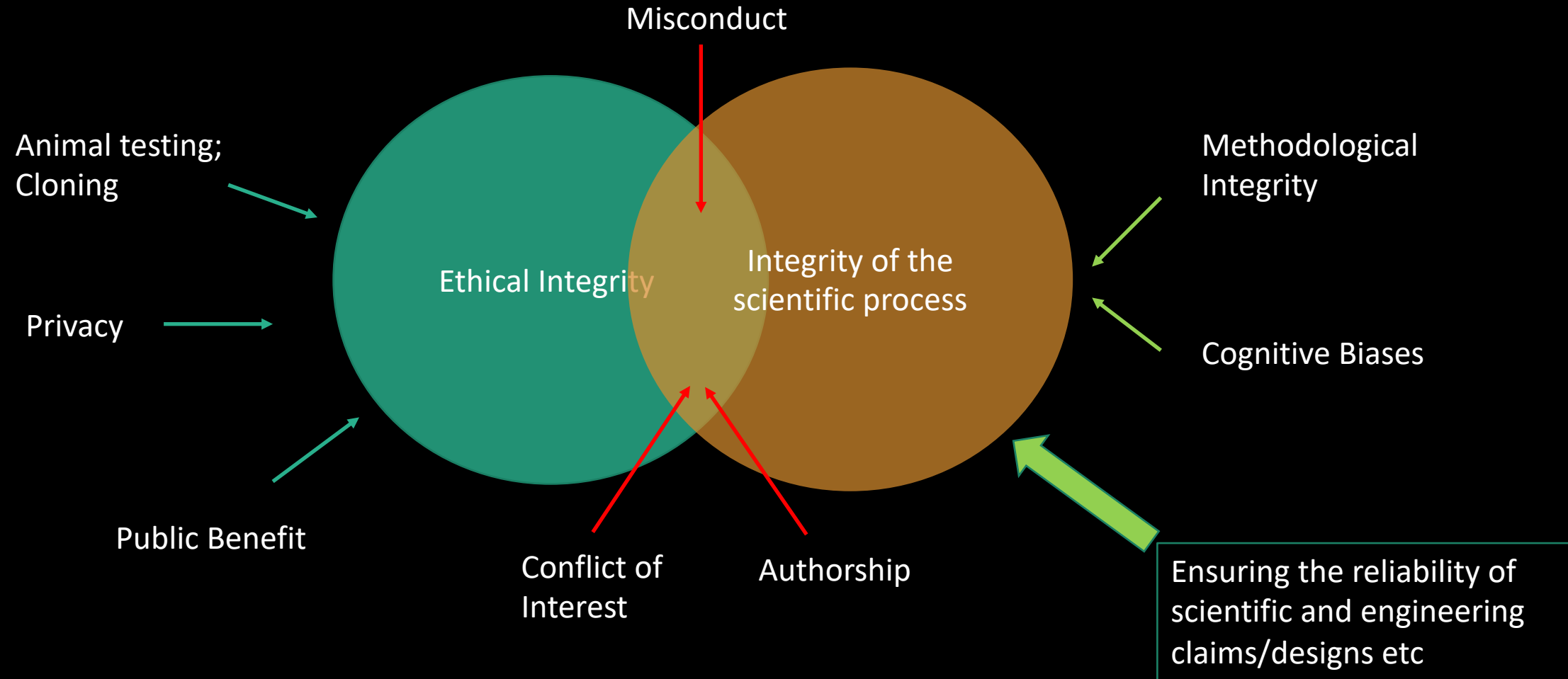Ethical Integrity

Integrity of the
scientific process

Components of Academic Integrity

Components of Academic Integrity

Components of Academic Integrity

Integrity of the Scientific Process

Norms of integrity

Universalism

Communalism

Disinterestedness/Impartiality/Independence

Organized skepticism

Robert K. Merton, 1942

Integrity of the scientific process

Integrity of the Scientific Process

Universalism

Communalism

Disinterestedness/Impartiality/Independence

Organized skepticism

Responsibility (accountability)

Recognition/Reward

Academic Freedom

Trust (in the system to produce reliable results)

Integrity of the scientific process

*Disinterestedness/Impartiality/Independence-> researchers have the right motivations + **cognitive independence***

Organized skepticism -> social-institutional mechanisms promoting debate, criticism and diversity (promoting cognitive independence)
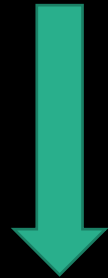
Methodological Integrity

What aspects of integrity does fabrication subvert?

What aspects of integrity does improper authorship subvert?

What aspects of integrity does p-hacking subvert?

Universalism
Communalism
Disinterestedness/Impartiality/Independence
Organized skepticism
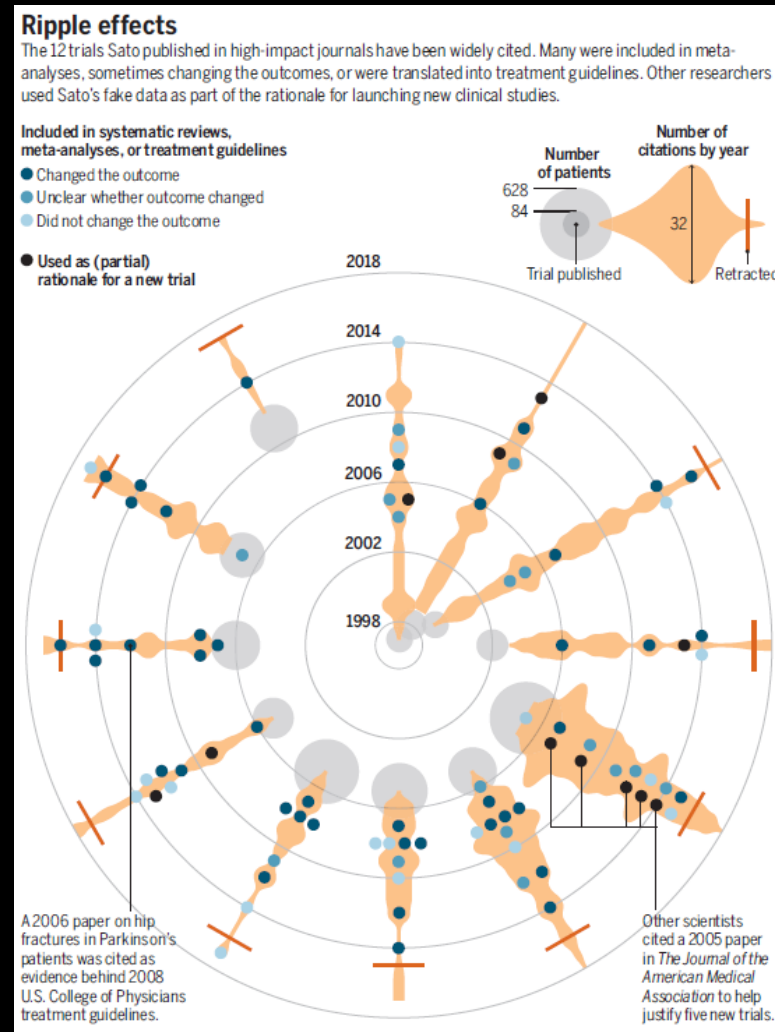*Academic Freedom*
Responsibility (accountability)
Recognition/Reward
Trust

Consequences of Fabrication

Ripple effects
The 12 trials Sato published in high-impact journals have been widely cited. Many were included in meta-analyses, sometimes changing the outcomes, or were translated into treatment guidelines. Other researchers used Sato's fake data as part of the rationale for launching new clinical studies.

Included in systematic reviews, meta-analyses, or treatment guidelines
● Changed the outcome
● Unclear whether outcome changed
● Did not change the outcome

● Used as (partial) rationale for a new trial

Number of patients
628
84
32

Number of citations by year
Trial published — Retracted

2018
2014
2010
2006
2002
1998

A 2006 paper on hip fractures in Parkinson's patients was cited as evidence behind 2008 U.S. College of Physicians treatment guidelines.

Other scientists cited a 2005 paper in *The Journal of the American Medical Association* to help justify five new trials.

"A Tide of Lies," by Kai Kupferschmidt, gives details in the August 17, 2018, issue of *Science*

Red Flags - Examples
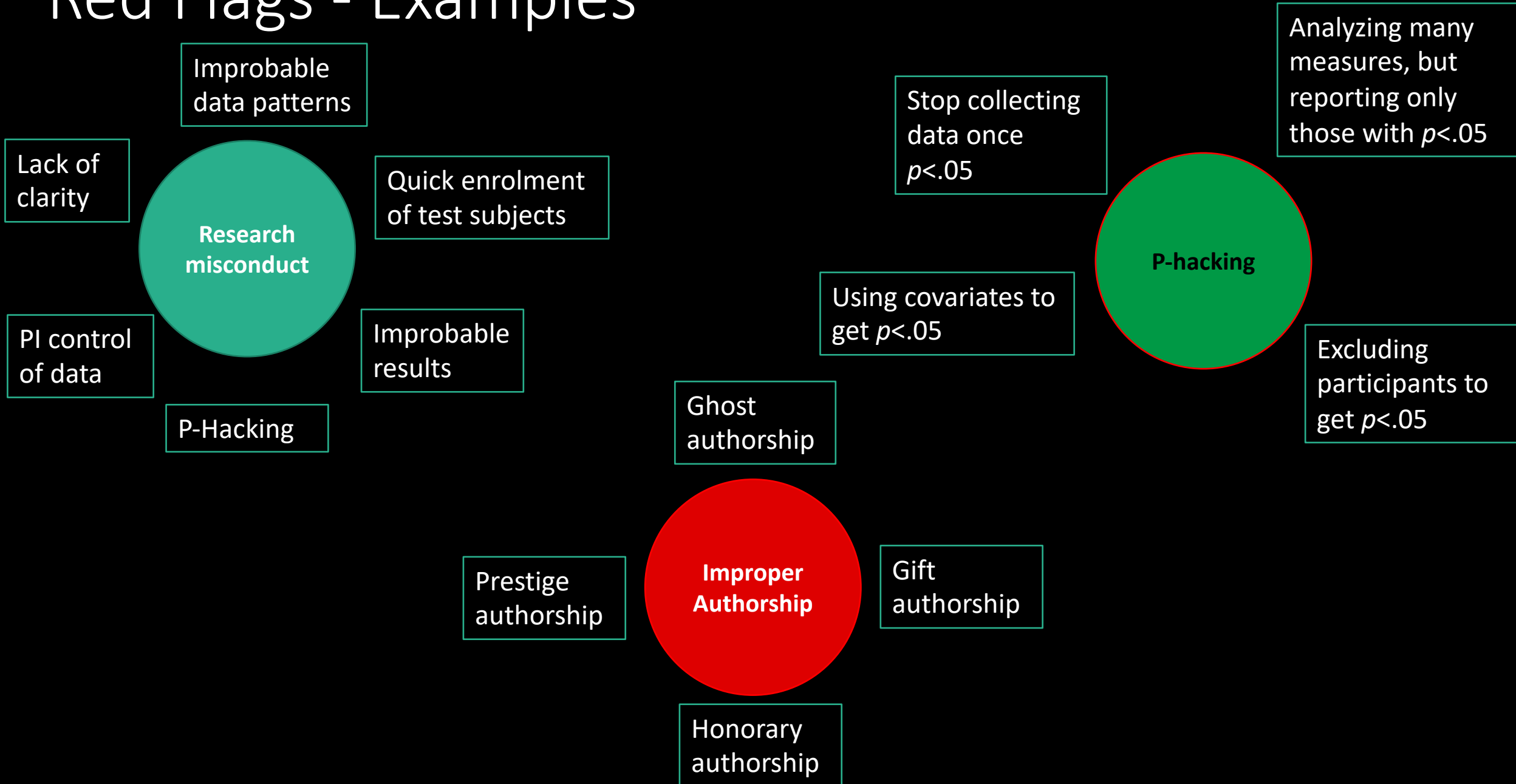
Focusing on Methodological Integrity

When do meth. integrity risks arise generally?

- Whenever a method, experimental paradigm, modeling framework, modeling idealization or abstraction, design assumption etc becomes so central to a practice that its original conditions of validity are "backgrounded".
- When such validity conditions are not well-explored in the first place.
- When outcomes are not independent of methodological choices and it's not a practice in the field to apply a "robustness" standard.
- When the limits of statistical and mathematical reasoning are not adequately understood.

Why might bad practices succeed?

Tractability

Institutional Reinforcement

Cognitive biases – such as availability

Example: Multiple Comparisons Problem
(Ioannidis, 2005)

Multiple comparisons increase the risk of a false positive.

Example: a researcher is interested in the relation between birth defects and parental features. p is set at 0.05;

H_1: birth defects related to mother's jobs

Gets p > 0.05…rejected

H_2: birth defects related to mother's age

Gets p > 0.05…rejected

H_3 : father job; …

H_n: {….weight, education, smoking, drinking…etc}

By the logic of significance testing the chance of at least one of n different tests producing a false positive is $[1 - (x)^n]$ where x = 0.95. This gets big quick.

(see also the Garden of the Forking Path problem – Gelman, 2013)

Example: p-hacking

- Manipulating study conditions to produce p-values under 0.05. It is most tempting when results are close to 0.05 anyway.

Six Ways to p-Hack (Leif D. Nelson)

1. Stop collecting data once $p<.05$

2. Analyze many measures, but report only those with $p<.05$.

3. Collect and analyze many conditions, but only report those with $p<.05$.

4. Use covariates to get $p<.05$.

5. Exclude participants to get $p<.05$.

6. Transform the data to get $p<.05$.

Other Poor/sloppy Statistical Practices

1. Poor interpretations of null hypothesis significance testing
2. Using stepwise regression to filter outliers
3. Best subsets regression
4. Overfitting
5. Univariate screening (data-cleaning)
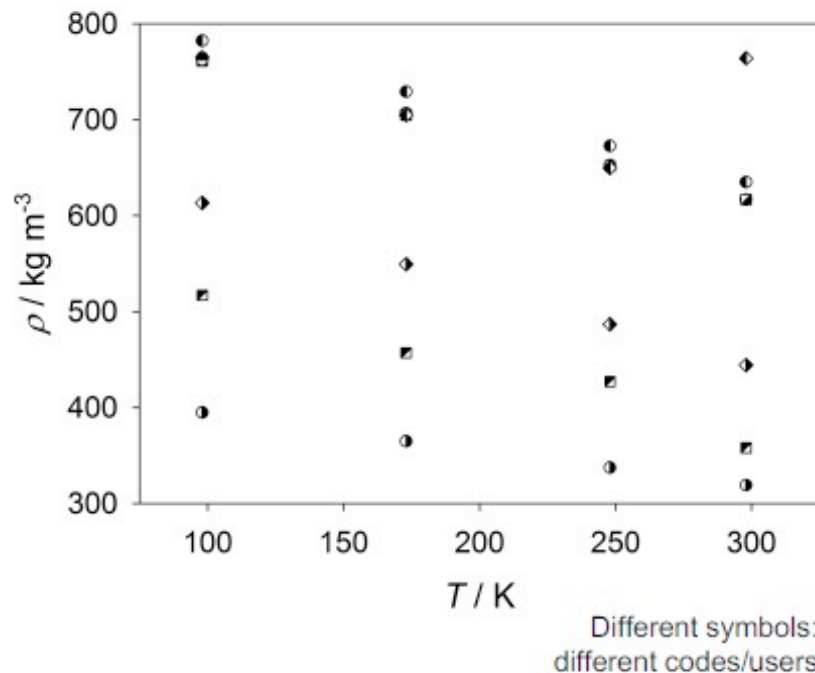6. Dichotomizing continuous variables

Common "Risky Assumptions" in Natural Sciences/Engineering

1. Re-applications of empirical models (outside their validity conditions)
2. Assumptions of linearity (which may be initially derived from nonlinear descriptions over certain conditions).
3. Treating environments as constant –> to turn relational properties into monadic properties and then drawing substantial inferences. e.g. "fitness" and "group selection" in population biology.
4. Treating systems under investigation as closed.
5. Treating fitting the data with a model as a primary success indicator; particularly if assumptions have been chosen to help in fact do this.
6. Failing to identify implicit methodological dependencies in modeling decisions.

Methodological dependency: Molecular dynamics Example


Results: n-Butane, Density @ 41 MPa, OPLS

After first iteration

Different symbols: different codes/users

Relative deviations

$$\frac{\Delta\rho}{\rho} = \left(\frac{Result\ of\ individual\ group - mean\ value\ of\ results\ of\ all\ groups}{Mean\ value\ of\ results\ of\ all\ groups}\right)_{Density}$$

Deviations between the results of the groups **larger than** statistical uncertainty of the data
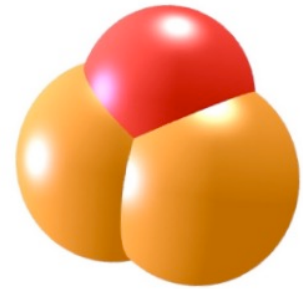
Molecular model (Force field)

Schappels et al. 2017
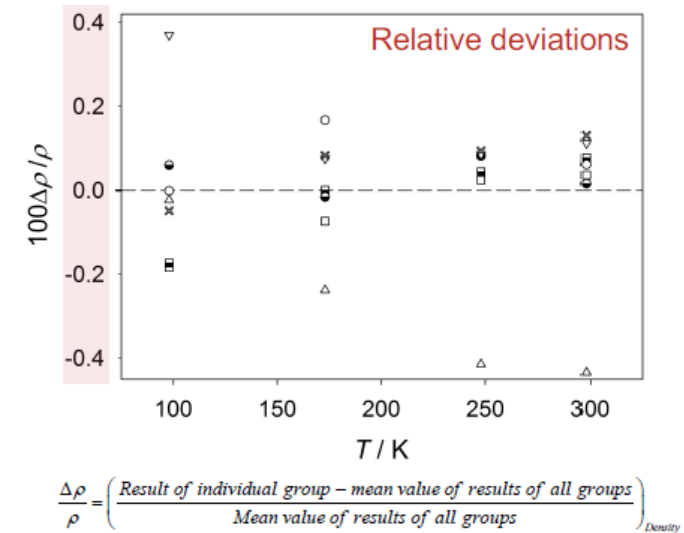
A crisis of reproducibility in the exact sciences?

Sources of Simulation Error

| Model x^{mod} | → | Coding | → | Implemen-tation | → | Simulation | → | Evaluation | → x^{sim} |

- Algorithms
  - Physical
  - Numerical
- Programming
- 3rd party codes

- Hardware
- Compiler
- Parallelization

- Control
  - Physical
  - Numerical
- Execution
  - Restrictions
  - Faults

- Algorithms
  - Physical
  - Numerical
- Programming
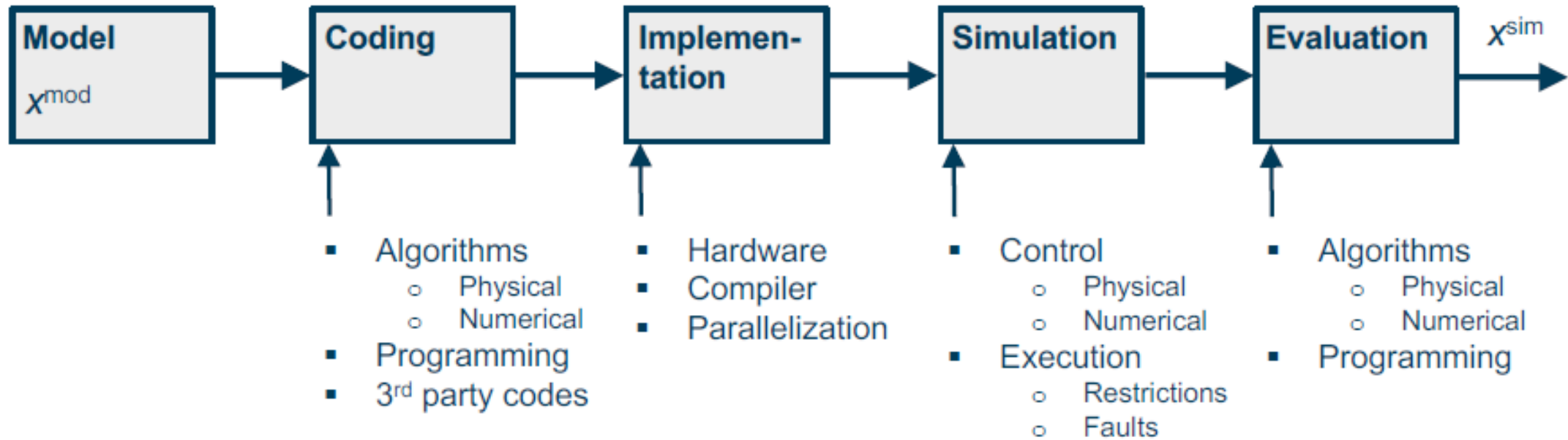
Common Cognitive biases

Confirmation bias

Selection biases

Congruence (direct testing) bias

Availability bias

Asymmetric attention bias

Consequences of Methodological Errors and Biases

- According to Ioannidis most published research findings are likely false. Ioannidis (2005). Why most published research findings are false. *PLoS medicine, 2*(8), e124.

- Current replication crises in the social and life sciences. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251)

- Careers in retrospect are threatened.

- Public health, social, economic and design decisions are being made on potentially dubious bases.

These put the overall <u>integrity</u> of academia under pressure.

How do you maintain methodological integrity?

-> Conceptual/cognitive resources for spotting biases/errors (cognitive independence – including awareness of common cognitive biases generally within a field))

-> Social-institutional mechanisms/practices promoting debate, criticism and diversity – and for providing these cognitive/conceptual resources.

Defending Methodological Integrity- ITC and elsewhere….

- Pre-registering methods (with journals)
- Replicability: independently repeating experiments (physical and simulation)
- Robustness: bringing to bear different types of evidence, and different independent methods.
- Publications of negative results.
- Publication of data or open access (NCCRI)
- Retraction
- Lastly, build cognitive independence….

NCCRI on Methodology

3.3.36. Be explicit about any relevant unreported data that has been collected in accordance with the research design and could support conclusions different from those reported.

3.3.38. Be explicit about uncertainties and contraindications, and do not draw unsubstantiated conclusions.

3.3.45. As far as possible, make research findings and research data public subsequent to completion of the research. If this is not possible, establish the valid reasons for this.

3.3.53 Be honest in public communication and clear about the limitations of the research and your own expertise. Only communicate to the general public about the research results if there is sufficient certainty about them.

To sum up...

- We all have a collective interest in maintaining scientific integrity.
- Scientific integrity requires more than simply good ethics.
- Scientific processes maintain their integrity due to institutional and cognitive norms and structures → but these may not be currently optimal.
- There is a real concern that things can go wrong particularly when the balance is wrong - > problems may not be readily perceptible.
- My goal through the TGS integrity programme and lectures like these is to provide some of these resources to students so they can raise their own questions and issues....

The Balance may be Wrong!

- It's not clear that current funding structures and reward structures promote integrity well.

- It's not clear that our correct overview mechanisms like peer review are well adapted to many aspects of integrity.

- It's not so clear that within our institutions we really promote or train cognitive independence well.