Lessons from the Past and Current Integrity Challenges

Klaas Sijtsma

Tilburg University

Paper presented at the Workshop on Enhancing Scientific Integrity: Progress and Opportunities in the Social and Behavioral Sciences

Washington DC, April 23-24, 2026

Three topics

1. Past cases of scientific fraud, including those that occurred at Tilburg University – characteristics, results
2. Lessons about oversight, research culture, and the systems needed to better detect and prevent misconduct
3. Integrity challenges presented by the advent of GenAI – examples, speculations, risks

Past case(s) of scientific fraud

The Diederik Stapel case

- Prior to September 2011 – Three PhD students in social psychology supervised by Zeelenberg wondered why PhD students supervised by Stapel obtained results so much better than they did:
  "Are they so smart or are we so dumb?"
  Started investigation checking results, documents, and so on
- August 2011 – Informed Zeelenberg about their inevitable conclusion: "Results reported in several articles must be fraudulent."
- Zeelenberg confronted Stapel, and then informed the Rector Magnificus (cf. provost), who confronted Stapel, who confessed; all in one weekend
- September 1, 2011: Rector informed me (vice dean of TSB) that Stapel (dean of TSB) was a fraudster; that he confessed; that he would be fired; and that they needed an interim dean (I) who would clean up the mess
- After I agreed, we installed a committee that would investigate the fraud
- September 7: Everybody involved was informed, then press release, show started and continued for the next 15+ months

Facts:

It was immediately clear that magnitude fraud and its implications were serious:

- Went back at least 15 years, covering periods at
  - ✓ *University of Amsterdam*: 1994—2000
  - ✓ *University of Groningen*: 2000—2006
  - ✓ *Tilburg University*: 2006—2011
- Involved dozens of articles and book chapters (90+), and affected several PhD dissertations (11)

Three committees were installed to formally investigate the fraud cases:

- For each publication it had to be secured whether it was based on falsified or fabricated data, or whether results were made up
  This was deemed necessary to
  - ✓ Inform colleagues and public about trustworthiness articles
  - ✓ Protect co-authors from being falsely accused of involvement
  - ✓ Safeguard former and present PhD students from their early careers being devastated
- They investigated whether the circumstances ("research culture") *possibly* facilitated the integrity breach, and
- They were asked to provide recommendations to prevent repetition

Results (presented in November 2012): The Committees found that

- The fraud committed inflicted ***great harm*** on coauthors, and ***PhD students*** in particular
- No one could be accused of ***culpable ignorance*** (i.e., no one knowingly cooperated in the fraud, people were misguided)
- A flawed performance of ***academic criticism*** facilitated—unintendedly—the long-lasting fraud; that is, nobody picked up available signs, such as
  - Unusual way of ***working in isolation***—closely together with PhD students, keeping senior colleagues at distance
  - ***Not*** allowing PhD students to collect their ***own data***—did not know was unusual because of isolation
  - Presenting ***unlikely results*** to journals—regular patterns, linear trends across conditions, huge effect sizes, etc.

FFP versus QRPs

Make a ***distinction*** between

Deliberate deception (**F**abrication, **F**alsification, **P**lagiarism)

(***rare***, bad for reputation science, sometimes damaging for people—for example, medical therapy based on fraudulent data can cause serious harm)

and

Unintentional erroneous practices (**Q**uestionable **R**esearch **P**ractices)

(not as visible but ***widespread***, resulting from lack of knowledge and experience in statistics and data analysis)

I believe that the main problem is not people like Stapel, but that we expect the ***research community*** can analyze their data all by themselves

Problem behind QRPs is:

- Researchers ***need to use*** methodology and statistics, but M&S are not their profession; they practice M&S ***on the side***
- Statistics is ***difficult*** and results are counter-intuitive; one is constantly misled
- Together with climate of performance pressure, the ***toxic mixture*** called Questionable Research Practices (QRPs) is unavoidable

Why is Statistics difficult?

Tversky & Kahneman (1970s) explained this:

Statistical reasoning requires ***rational and conscious reasoning***, costs a lot of effort, people do not do this; first inclination is to follow ones ***intuition*** and jump to conclusions

Intuition works two ways:

- Based on ***experience***; makes you do approximately the right things, before you have started thinking about a well-founded answer

- Based on ***heuristics***; makes you respond using cognitive automatisms that replace the difficult but correct question (impossible to answer) with an easier but incorrect question (answerable); puts researcher on wrong track

Third reaction is cognitively demanding: Try to solve problem ***rationally***;

this is ***not*** what we do first, we rather react intuitively

Example difficult question (not for you, I guess)

Select 23 persons at random; what is the probability that at least two have their birthday on the same date (except 29 February)?

Intuitive, ***experienced*** response: Is combinatorial problem, be careful, such problems are deceptive. *Take your time!*

Intuitive, ***heuristic*** response: Replace with simpler question: “Do I know people who have their birthday on the same day?

Answer: No, so, probability must be small

Demanding, ***rational*** answer (Solve the problem neatly; nobody does this):

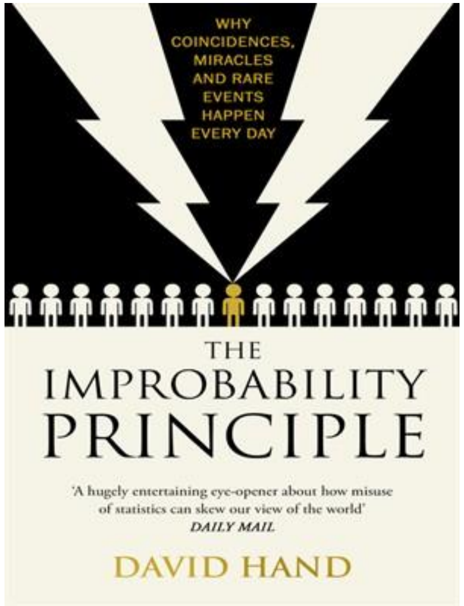
$$P > \mathbf{.5}$$

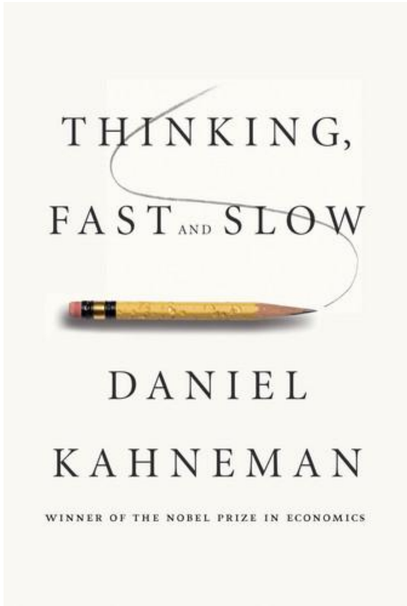
Notice: Intuitively, one ***expects*** a small probability, but the probability is large, and people do not understand why; this is what I mean, intuition (based on *heuristics*) fails

Further reading (let others convince you):

Hand, D. (2014). *The improbability principle. Why coincidences, miracles and rare events happen every day*. London, UK: Penguin Random House.

Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin Books Ltd.





Lessons Learned: *Policy* instead of Statistics

Conclusion: Statistics is difficult for ***everybody***, including statisticians

Does additional statistics ***education*** help (to prevent QRPs)?

- Statistics education shows researchers what's in the toolbox, and what you can use the tools for; ***necessary knowledge***
- More education provides more methods, **no *experience***

So, what to do?

- Help people to ***recognize*** the method that suits the problem at hand
- Teach people to recognize situations ***in which they need help***
- If you do, stop suggesting ***everybody*** can do statistics
- Be careful teaching "***crash courses***" in multilevel analysis, etc.

Here are three ***policy recommendations*** I prefer:

I. Require researchers to (pre)register their research

Main benefit for fostering **R**esponsible **R**esearch **P**ractices is a heightened awareness of

- the importance of ***thinking*** before you execute the research, and
- the danger of ***tampering*** with the hypotheses, the methods, and the data until they confess to what you need to know

Disclaimer:

Beware of methodological ***rigidity***; it would imply throwing away all research (weak theoretical underpinnings, noisy data)

Emons, W. H. M., Sijtsma, K., & Bouter, L. M. (2025). Registration of research on research integrity is still not common: Findings from the Hong Kong, Cape Town, and Athens editions of the World Conference on Research Integrity. *Accountability in Research: Ethics, Integrity and Policy,* 33:4, 2575442. https://doi.org/10.1080/08989621.2025.2575442

II. Make Data Publicly Available

Wicherts et al. (2006): 73% of the authors of APA journals did not provide their data upon request

Researchers are quite reluctant, because they don’t want to be scooped

- Allow researchers first use of data until they have published
- “Owner” of data can be coauthor with others using their data
- Data sharing produces new results and develops networks
- Researchers must negotiate with owner data (bank, hospital); data must be public property, results must be controllable

III. Involve a Methodologist or Statistician in Data Analysis

Hire intuition based on ***experienc*e** that you do not have yourself

- Simmons, Nelson, & Simonsohn (2011, *PsychScience*): Understanding and use of statistics are ***difficult*** and are the main causes of errors made in data analysis

- Tversky & Kahneman (1970s): Explanation is absence of ***experience***, thus use of wrong intuition based on ***heuristics***, causing researchers to fall into all the traps set by counter-intuitive statistics

A Controversial Recommendation

Responding to the Stapel data fraud affair in 2011, Tilburg School of Social and Behavioral Sciences installed the ***Science Committee***

Science Committee—What is it, what does it do, and why?

- *Audit* committee samples 20 of approx. 500 articles from School's data base; assesses ***quality of data storage*** and ***reporting of research methods***
- ***Advises*** researchers about data storage, completeness data sets, honoring subjects' privacy, access to data, and data availability
- Aims:
  - ✓Encourage concerted effort to improve ***accountability*** for data handling and methods reporting
  - ✓Create opportunity for all to ***learn***; ***not a witch hunt***
  - ✓Contribute to development **university's data policy** and to a *Dutch* ***national protocol*** concerning data archiving by researchers in social and behavioral sciences

Science Committee was installed in 2012, and started working as follows

- Set up rules and regulations for researchers' data handling
- Announced annual random audits (20 of 500 articles)
- Research groups devised their own data (storage) policy that suited their needs best; initially no general policy

Works quite well, but not perfectly:

- Some groups have their data policy better in place than others
- People tend to arrange their data storage only when they are audited
- When people have left the School, they tend to loosen commitment
- No consistent policy for universal data storage system, underway
- There remains much work to do, but creates greater awareness, stronger sense of responsibility and accountability

Labib, K., Tijdink, J., Sijtsma, K., Bouter, L., Evans, N., & Widdershoven, G. (2023). How to combine rules and commitment in fostering research integrity? *Accountability in Research: Ethics, Integrity and Policy*, *31*, 917-943. https://doi.org/10.1080/08989621.2023.2191192

GenAI: Blessing or Disaster for Research Integrity?

Examples of GenAI "intrusion" in research:

- Editor *Psychometrika* to associate editors (March 13, 2026):
    - Rejected manuscript because of typical AI features
    - Reviewer declined a review request because of suspicion of AI use
    - Publisher: Decline based on quality, not AI suspicion (legal consequences)

- Read data: Program communicates via dialogue box and does a statistical analysis, including generating R-code, tables and figures. Other programs generate article with your name on it. Can be done in minutes.

- Colleague sent me a paper he found in *Research Gate* with his name on it; it was not his paper but probably generated with AI; why?
  Proved difficult to have it removed from Research Gate

Cognitive Automation in SAP Supply Chain Modules: Designing Predictive Disruption-Response Systems

Dr. Wil van der Aalst
Department of Mathematics and Computer Science
RWTH Aachen University, Germany

Prof. Jan vom Brocke
Institute of Information Systems
University of Liechtenstein, Vaduz, Liechtenstein

Publication Date: December 2025

Abstract

Global supply chains have become increasingly vulnerable to geopolitical instability, climate-induced disruptions, demand volatility, and supplier concentration risks. Traditional SAP supply chain modules provide transactional visibility but remain largely reactive in responding to operational disturbances. This study proposes a cognitive automation framework embedded within SAP S/4HANA Supply Chain modules to design predictive disruption-response systems capable of anticipatory decision-making. Leveraging adaptive machine learning algorithms, process mining techniques, and real-time in-memory analytics, the research conceptualizes an integrated architecture that detects anomaly signals, forecasts disruption probabilities, and autonomously recommends mitigation strategies. A quantitative quasi-experimental design is adopted to compare conventional rule-based SAP workflows with cognitive automation-enhanced systems using simulated disruption datasets. The study contributes to enterprise systems scholarship by bridging cognitive AI theory with operational supply chain resilience within ERP ecosystems.

Keywords: Cognitive automation, SAP supply chain, predictive disruption management, ERP intelligence, process mining, adaptive AI, supply chain resilience

---

1. Introduction

Supply chains operate within increasingly complex and interconnected global networks. Events such as pandemics, trade restrictions, cyber threats, and raw material shortages have revealed

Stapel and all other fraudsters had to work hard to deceive their colleagues in a believable way; with GenAI, they could have been much more ***efficient*** and ***credible***

Credible?

Stapel was not good at statistics; he did not understand probability and suffered from all the intuition flaws, and his colleagues suffered from the same problems

Thus, he was hard to catch by his colleagues

What ***worries*** me most

Does the availability of GenAI lower the threshold for fraud – FFP?

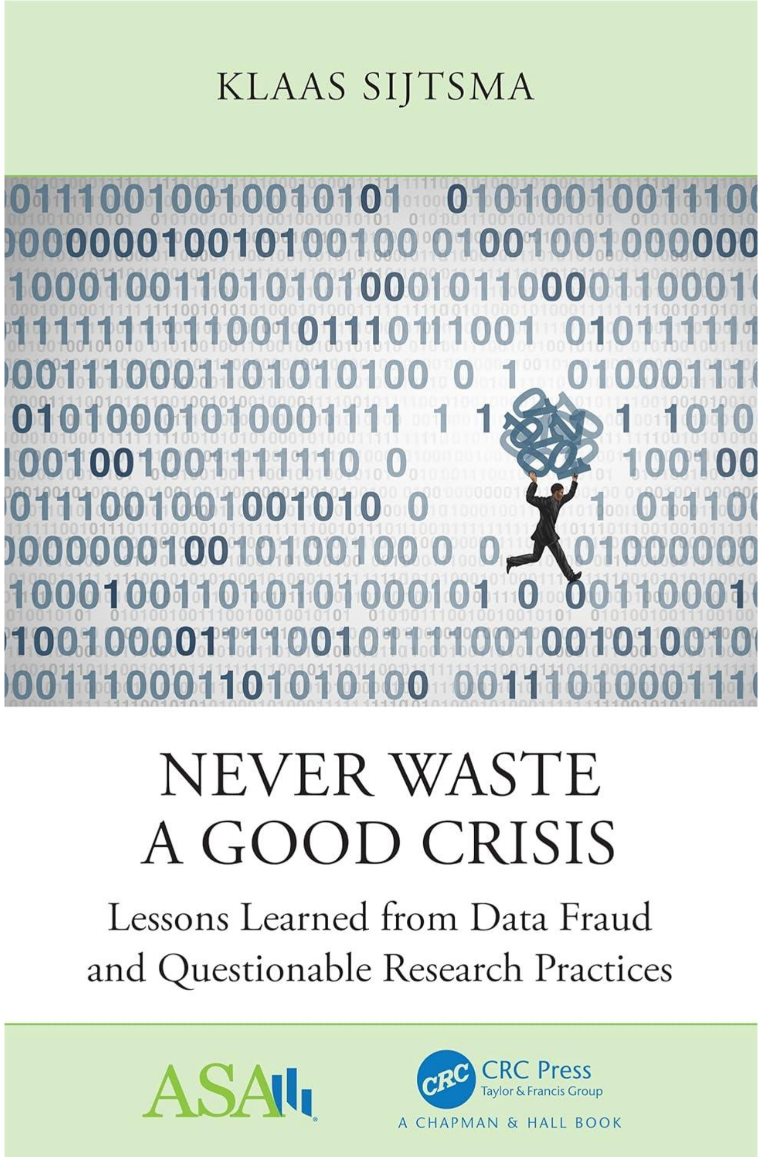
Recent surveys: 65-83% of all students say they use GenAI

frequently

and with

many study activities (e.g., preparation courses, summarizing literature, executing assignments)

What do you think?



Sijtsma, K. (2023). *Never waste a good crisis. Lessons learned from data fraud and questionable research practices*. Boca Raton, FL: Chapman & Hall/CRC

Additional references:

Sijtsma, K. (2016). Playing with data—Or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika, 81*, 1-15.

Labib, K., Tijdink, J., Sijtsma, K., Bouter, L., Evans, N., & Widdershoven, G. (2023). How to combine rules and commitment in fostering research integrity? *Accountability in Research*. https://doi.org/10.1080/08989621.2023.2191192

Emons, W. H. M., Sijtsma, K., & Bouter, L. M. (2025). Registration of research on research integrity is still not common: Findings from the Hong Kong, Cape Town, and Athens editions of the World Conference on Research Integrity. *Accountability in Research*. https://doi.org/10.1080/08989621.2025.2575442

Thank You

?