

Contents

The Sign Flip Error	2
A story shared by Isaac Yonemoto.....	5
Under Pressure	7
Who Is An Author?	10

The Sign Flip Error

The real life case about a group of Danish physicist who have a sign flip error in their algorithm, fail to share their code, and listen to some bad advice.

A group of physicist lead by Professor Mads Jacob Sherson and Professor Klaus Mølmer from Aarhus University, Denmark, publish a high profile paper in Nature [1] in April 2016, showing that “human players succeed where computer algorithms fail” when solving gamified optimization problems in quantum physics. The findings are reported in several science and popular media outlets celebrating that human intuition can “outperform the best quantum algorithms” and go “beyond the classical laws of physics” [2]. Sherson receives numerous accolades, including a DKK 3.7 million grant from the Carlsberg Foundation.

Assistant Professor Dries Sels from New York University doubts the result and in July 2017 asks for a copy of the code, a request Sherson and Mølmer ignore. Frustrated, Sels performs an independent replication in September 2017 [3], which he later publishes in Physical Review A [4], showing that another decade-old algorithm works much better than any player solution, making the claims by Sherson and Mølmer “untenable”. Sherson and Mølmer find the paper confrontational and decide to ignore it. Soon after colleagues at the department start to raise concerns about the missing response and talk about breaches of research integrity. Sherson and Mølmer seek advice from an internal advisor for responsible conduct of research as well as from the Head of the Department. They agree that they see no reason start a “mud-throwing” public discussion with a “young researcher” who, according to Sherson “at the time did not attract any attention apart from Aarhus University” [5].

Question 1: Is it appropriate to ignore a request for codes and scripts, even if the request comes from a young and unknown researcher?

Question 2: Is it appropriate to try to replicate the results of a paper that is published in a high impact journal such as Nature?

Question 3: What are the standard practices for sharing data, metadata, code, protocols etc. in your research discipline? Are they typically shared with others? If so, at what time in the project, how and with whom? If not, why not?

Internal critique keeps getting louder. Three colleagues at Sherson’s and Mølmer’s department argue in an internal discussion board “that good scientific practice must be that such professional disagreement is conducted as an article/comment and exposed to peer review”. Computer scientist Allan Grønlund, who is employed as a postdoc at the same university as Sherson and Mølmer, joins the discussion in April 2019 by showing that other basic algorithms, including the one used by Sherson and Mølmer themselves, easily outperforms all player solutions [6], effectively confirming Sels’ conclusion. Like Sels, Grønlund makes his code and analysis publicly available on GitHub and ask the authors of the original nature-paper to do the same in order to find the reason behind the large difference in the results.

Internal meetings at the department try to calm down increasingly agitated colleagues, but the leadership and their advisors stand behind the researchers of the Nature article arguing that nothing out of the ordinary has happened. As a sign of compromise, Sherson and Mølmer are encouraged by the Head of Department to answer Sels and Grønlund. But instead of providing any code or explanation, Sherson and Mølmer publish a comment [7] in October 2019, disputing the relevance of the replications made by Sels and Grønlund as being the result of an “advanced arsenal of

algorithms” that misses the point made in their original paper, which was not so much that humans can ‘outperform’ computers, but only that “player solutions can be employed either as seeds for numerical optimization algorithms or as guidance towards low-dimensional representations of the solutions”.

Question 4: How did Sherson, Mølmer, and the leadership handle the growing conflict? Any suggestions for improvement?

Question 5: What could Sels and Grønlund have done in order to receive the code much earlier?

In December 2019, after having waited for nearly six months, Grønlund finally receives the (MATLAB) script used by Sherson and Mølmer. It turns out that it has a simple sign flip error, which essentially means that the algorithm tries to find the worst possible solution instead of the best. Any basic analysis of the output of their algorithm would reveal this obvious flaw in the implementation [8], and when the error is corrected, the results turn out to be essentially the same as the results in all the algorithms tested by Sels and Grønlund.

Question 6: Sherson and Mølmer would have benefitted from a quality control of their code and data prior to publishing their results. What quality control measures do you take in your research, so that you can trust the code and data that form the basis of your results?

Sherson, Mølmer, and the Head of Department are informed about the error. Three months later, in February 2020, the rest of the department, the journal Nature and the Carlsberg Foundation are informed as well. In May 2020 Sherson and Mølmer are found guilty of questionable research practice by the Practice Committee at Aarhus University for not providing the code in a timely fashion. Nature retracts the paper in July 2020.

This should have been the end of the story. It turns out, however, that during the whole process, the influential fund manager at The Carlsberg Foundation, Flemming Besenbacher, who gave the grant to Sherson and who also is employed at Aarhus University, tried to influence the process by sending highly agitated emails to Sherson, Mølmer, and the leadership. The emails are leaked to a major Danish newspaper, showing that Besenbacher attacked the internal critics as being “disloyal”, “baboons”, who “shit in their own nest” and should get “whacked on the head” or “there will be consequences”. Even the Dean of the Faculty of Science, Karsten Pedersen, joins the pummeling party in November 2020 by writing a private email to Besenbacher: “I’m about to put duct tape on the mouth of many people and tie their hands on their backs until they have thought about what they are doing.” [9]

A shitstorm breaks out in Danish media and Besenbacher and Pedersen are forced to apologize for their inappropriate choice of words. The damage control is too little and too late, however. 26 other researchers from Aarhus University write in the spring of 2021 a letter to their union representative expressing concern about their leadership being unduly influenced by external stakeholders and not being able to handle a simple instance of questionable research practice in a neutral and timely fashion. In addition, a claim emerges that there exists an old version of the code with a timestamp from May 2017, which does not have the fatal sign flip error. Once more, Sherson and Mølmer are reported for questionable research practice, which is awaiting a decision from the Practice Committee. *(to be continued...)*

Question 7: Clearly, communication has gone terribly wrong in this case, and a culture of mistrust has infected parts of the university. If Sherson and Mølmer had responded to the

request by Sels and acted quickly, much of this could (maybe) have been avoided. In your own experience, have you ever experienced a conflict between your research integrity and your loyalty towards your colleagues, bosses and/or collaborators? How did it go?

Question 8: The existence of an old version of the code suggests that multiple versions exist. We can speculate that Sherson and Mølmer somehow ended up using wrong versions. What is your strategy to versioning, file naming and research documentation? How do you keep track of changes made during the research project, in particular when working with multiple collaborators on code, or to collect data?

Question 9: The journal Nature has a detailed policy on data, code and materials availability, which requires authors to make materials, data, code, and associated protocols promptly available to readers without undue qualifications [10]. In 2014, the Nature journals also adopted a “code availability” policy to ensure that all studies using custom code deemed central to the conclusions include a statement indicating whether and how the code can be accessed, and explain any restrictions to access. [11] Since September 2016, authors are required to make a data availability statement, and since 2018, authors are required to fill out a Code and Software submission checklist [12]. Which of these publisher requirements did Sherson and Mølmer breach?

Question 10: Have you ever tried to reproduce someone else’s research? Have you ever tried to incorporate code or data sets produced elsewhere in your research? If so, what were your experiences? (How) could we improve the reproducibility of research projects and the reusability of code and data?

1. Sørensen, J. J. W., Pedersen, M. K., Munch, M., Haikka, P., Jensen, J. H., Planke, T., ... & Sherson, J. F. (2016). Exploring the quantum speed limit with computer games. *Nature*, 532(7598), 210-213.
2. <https://www.nature.com/articles/532184a>
3. Sels, D. (2017). Can math beat gamers in Quantum Moves? <https://arxiv.org/abs/1709.08766>.
4. Sels, D. (2018). Stochastic gradient ascent outperforms gamers in the quantum moves game. *Physical Review A*, 97(4), 040302.
5. <https://videnskab.dk/forskerzonen/kultur-samfund/efter-afsloering-af-strid-mellem-forskere-skarp-kritik-er-afgoerende-for-at-sikre>
6. Grønlund, A. (2019) Algorithms Clearly Beat Gamers at Quantum Moves. A Verification, <http://arxiv.org/abs/1904.01008>.
7. Mølmer, K., & Sherson, J. (2019). (Of course) algorithms beat gamers at solving quantum control problems. <https://arxiv.org/abs/1910.14364v1>.
8. Grønlund, A. (2020). Explaining the poor performance of the KASS algorithm implementation. <https://arxiv.org/pdf/2003.05808.pdf>.
9. <https://www.berlingske.dk/samfund/kritik-af-forskerfejl-vakte-dyb-vrede-kritikere-blev-kaldt-redeskidere-og>.
10. <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#data>
11. <https://www.nature.com/articles/d41586-018-02741-4>.
12. <https://www.nature.com/documents/GuidelinesCodePublication.pdf>

A story shared by Isaac Yonemoto

A real life case about suspected misconduct and smart career moves, shared online and having generated several similar stories in the comments.

"In graduate school, in my lab there was a grad student who was kind of an unlikely 'professor's pet'. He was tall and had surfer's long hair with a bit of a hippie aesthetic. Anyways, he was also really completely clueless about how to do science correctly, but also, I guess, really good about playing politics (there was a time when he asked me to put some bacterial plasmid DNA on my mammalian cells. I told him "it doesn't work that way", but I did it anyways and handed over the cells, and he got the observation he was expecting).

On his main project he was teamed up with a super sketchy foreign postdoc that I was convinced would say anything to get high profile papers out. So they did a series of experiments and reported results that screamed "artefact". On one of them, for example, the postdoc got trained to use the electron microscope and they went through thousands and thousands of images to pick out the one that had "just the right morphology" (I am pretty sure they were snapping photos of salt crystals). On another, they reported that their research subject protein was so fast at the process we were studying that everything occurred IN MIXING TIME. That to me, screams "you are not doing your experiments carefully". Meanwhile I was sweating balls working on a very careful preparation of similarly finicky proteins (you agitate them and they do bad things since they're metastable) and finally got it to produce reproducible results. I suggested they adapt my preparation to their protein but they couldn't give a damn, they had already published their paper and had moved on to sexier proteins."

Question 1: Did you yourself ever have similar suspicions about a co-worker?

Question 2: Do you think Yonemoto should have made his suspicions known? How?

Question 3: In fact, we do not know if Isaac Yonemoto was a fellow graduate student at the time, a lab instructor or the supervisor of the student mentioned. Would his options and obligations be any different if he were one or the other?

"But then an intern was put on the project, and she could not reproduce their results, after working on it for six months (she is careful and honest). At the end, I felt so bad for her, I offered to train her on my technique, but she passed. I think she was burned out on the project. I asked if I could get a sample of the protein that she had prepped, and she agreed. I ran the protein through my preparatory technique and observed that there was a contamination that could have seeded the kinetics of their process. Upon isolating an uncontaminated sample, I carefully but briskly rushed the sample over to the machine. Nothing. Curious, I jacked the temperature up to get it going faster. Nothing. I left it in the machine overnight. Nothing. Finally, convinced that I had likely done something wrong, I dropped the sample in a shaker at temperature, came back the next day and recorded amazingly high signal. In short, the observation that it was "super fast" was entirely an artefact.

As I, too, was trained on the Electron Microscope, I quickly spotted my sample onto an EM disc, reserved some time and hopped on the 'scope. The first grid sector I looked at, there was literally TEXTBOOK morphology in front of my eyes. I stapled together my results, gave it to the grad student, and told him that the general gist of his paper was probably still correct, but that he should be careful about characterizing his protein as exceptional. I then said it was in his hands to do the right

thing. What do you think he did? Nothing, of course. He kept on the talks circuit, still talking about how exceptional his discovery was, and to date there have been no retractions. He even won the NIH grad student of the year award.”

Question 4: Are there other things Yonemoto could have done apart from telling the grad students about his results and that he now “should do the right thing”?

“The epilog is that after a decade of floundering I realized that even though I am pretty good at science, I was no good at playing academic politics and quit the pursuit; I drove for lyft/uber for a bit, and now I'm a backend dev. I am certain that my experiences are not unique. Amazingly, the intern returned to our lab, and had her own three-year stint chasing ghosts that turned out to be overoptimistic interpretation of results reported by a postdoc. Oh. What happened to the grad student? He's a professor in the genomics department at UW.”

Question 5: Do you think it would be possible to find out who is the ‘professor’s pet’ using the information above?

Question 6: Have a look on the comments of the original post (link in the reference) and try to read a few of them. What strikes you the most? What does that tell you about the commenters? About the media? About the state of scientific practice?

References: <https://news.ycombinator.com/item?id=25922799>

Under Pressure

A case about publication pressures, competition, favoritism, and unclear expectations.

Sam Best is a post-doctoral fellow who has worked in Assistant Professor Taylor Jones's lab for almost 5 years. Best is now working on a project investigating how cells respond to a particular stimulant. Jones is a Tenure-Track Investigator coming up for tenure consideration within the year, who established the cell stimulation response system upon arrival at NIH, but Best later modified and perfected it. Best reported the development of the system and proof-of-principle data in two peer-reviewed publications, including one as first author. Their new research showed that the cellular response to the stimulant *Invigorin* was initially low but then steadily increased over time, accompanied by expression of a particular protein within a subset of the cells. Best found that adding specific chemicals inhibiting expression of that protein eliminated the cellular response. Best and Jones conclude that the protein mediates the effect and that they have uncovered a novel mechanism by which cells respond to this class of stimulants.

They draft the manuscript and send it to a high-impact journal. Jones believes their findings represent a major advance that could increase the likelihood of achieving tenure. The journal responds that while reviewers believed the work is exciting and potentially impactful, they want more direct evidence to prove the model through additional experiments, implying that the paper will be accepted if the new experimental data support the model.

Question 1: What kind of message do reviewers send when they ask for evidence to 'prove' a model? What are the pitfalls of trying to 'prove' a hypothesis?

Meanwhile, Best is reaching the end of his NIH appointment and begins a geographically restricted job search in an effort to join his partner, who had moved for a job months earlier. Luckily, Best receives an interview invitation from **Innovative Pharma**, a prestigious company in the targeted area. Best also makes the short list of applicants for a position at **World's Fabulous Research Institute**, which provides opportunities for exciting scientific collaborations. The institute position is a dream job but requires preparing for a research proposal and an in-person interview within the next few weeks. Because the institute job is the first choice, Best delays the pharmaceutical company interview process until hearing from the research institute, even though the company position has a higher salary and is an excellent backup option.

Jones really wants to complete the reviewers' suggested experiments quickly and publish the study because it would increase the potential of achieving tenure, but Best is concerned about not being able to finish the work while applying for the institute position. Best relays these concerns to Jones and suggests that they ask Kai Ettero-Sanson, a new post-doctoral fellow that Best trained over the past year, to conduct the experiments, saying Ettero-Sanson would be eager to work on the project. However, Jones asserts that Ettero-Sanson needs more experience because the system is 'finicky' and implies that Ettero-Sanson has lesser lab skills because of training outside the United States. Jones tells Best not to worry because even if neither position comes through, more offers will come, and compliments Best again for being "very gifted at the bench," a comment Jones has made many times. Jones adds that Best will be able to stay at NIH for an additional sixth year without a problem and that having a first-author paper in the **Journal of Fantastic Results** will greatly improve job prospects.

Question 2: Is it fair to ask Ettero-Sanson to become involved with the project at this point? What are the advantages/disadvantages of having another researcher perform these experiments?

Question 3: Is the advice from Jones about Best's job search reasonable? What would prompt Jones to offer this advice?

Question 4: How should a lab handle systems that tend to be 'finicky'; i.e., a system that is reliable but requires extremely strict adherence to the protocol?

Question 5: Do you think Jones has a bias against Ettero-Sanson? How could a bias (or the perception of one) affect lab relationships, pressure, and career development?

Best reluctantly agrees to ask the institute to postpone the in-person visit and convinces Jones to allow Ettero-Sanson to help with the experiments. It takes weeks for Best and Ettero-Sanson to finish their work, but the results are confusing and in one case contradictory to what they predicted. Best shows the data to Jones, who concludes that the results must be incorrect and that perhaps Ettero-Sanson had misread reagent bottles or protocols. Jones suggests that Best repeat the experiments, but Best reminds Jones that the institute has been trying repeatedly to schedule the on-site interview ASAP. Jones then asks: "Do you think this institute position is a good fit for you? I say this because it is a very competitive environment, and I've found that success in places like that depend on one's ability to think broadly and develop novel and creative ideas." Best is troubled by these remarks because they imply that Best might not succeed as an independent scientist. It reminded Best of a previous comment by Jones that fellows who received PhDs from "certain types of universities" are typically better suited for non-academic positions. Best also realizes that aside from repeated compliments on technical skills, Jones has never commented on Best's potential to be a PI/group leader or suggested additional training or experience that would help with achieving a leadership position. Best is now worried about the recommendation letter that Jones had written, what had been communicated privately to professional colleagues, and whether successfully completing revisions of the paper would affect future letters.

Question 6: How might mentoring/communicating be improved in this interaction?

Question 7: What should take place during a conversation in which a trainee asks their PI/group leader for a letter of recommendation? What is the role of the PI/group leader in that conversation?

Best works day and night, mostly alone in the lab, repeating the experiments and finishes them faster than any of the previous experiments. This time, the data trended as expected. Jones is happy and immediately encourages Best to write up the results without Ettero-Sanson as a co-author and to resubmit the paper, commenting how this will help both of their careers. Best is relieved. While both potential job opportunities had granted interview delays, they were clear that no further delays would be acceptable.

Question 8: Is it proper to remove Ettero-Sanson as an author? How and when should Jones have communicated how authorship on this paper would be decided?

Question 9: Is running experiments 'day and night' appropriate in this case? What issues can arise from this behavior?

Ettero-Sanson learns of the new results and is skeptical. A meticulous experimentalist, Ettero-Sanson does not believe the new results could differ so substantially from the data obtained together with

Best. After learning about the change in authorship, Ettero-Sanson tries to move on but cannot and decides to investigate further. One day, after everyone has left the lab, Ettero-Sanson looks through Best's lab notebooks and electronic files and uses the Excel data to try to replicate the results, without realizing that doing so would destroy the integrity of the spreadsheet. From the analysis, Ettero-Sanson concludes that Best ran the most recent experiment multiple times but presented only results from the three best experiments to Jones.

Question 10: Is Ettero-Sanson justified to suspect Best's results? If so, what should Ettero-Sanson do?

Question 11: Why is the integrity of primary data so important? How can the integrity of computer files be maintained?

Question 12: Is it acceptable to present data selectively? Under what conditions, if ever, can specific data sets be removed from an analysis?

Ettero-Sanson is worried about the consequences of coming forward and questioning the experimental results, but out of great concern, speaks with Jones about the possible misconduct. Jones brushes off the concerns, saying that Ettero-Sanson must be mistaken and implies that Ettero-Sanson misunderstood Best's lab notebook and files, perhaps because of language issues. Jones begrudgingly agrees to a formal meeting to discuss the issue further but neglects to schedule one. Ignored and upset, Ettero-Sanson contacts the NIH Agency Intramural Research Integrity Officer (AIRIO). A preliminary assessment indicates that a misconduct inquiry is warranted.

Question 13: How should Jones respond to Ettero-Sanson's concerns?

Question 14: What type of signals is Jones sending to Ettero-Sanson by bringing up 'language issues' and by not scheduling the meeting?

During the misconduct inquiry, Jones worries that rumors will spread, required external reference letters will be tainted, and the tenure committee will not recommend promotion. Jones blames Ettero-Sanson for the entire situation and begins to wonder if another lab would be a better fit. Ettero-Sanson worries that relationships within the lab are irreparably harmed. Best is extremely distressed and concerned about reputational damage. Unable to concentrate on the job proposal, Best withdraws from consideration for the institute position, but does interview with the pharmaceutical company as the inquiry progresses.

The inquiry ends and concludes that no further investigation is practical because Ettero-Sanson's handling of the original Excel file compromised its integrity. The pharmaceutical company selects a different candidate, and when Best asks for feedback, the recruiter responds that Best seemed distracted during the interview.

Question 15: What choices could have been made differently that would have led to positive outcomes for everyone in this case?

Question 16: Have you ever encountered or heard about any other situations related to the themes of this case study?

Credits: This text is adapted for use in Denmark from the case study "Science Under Pressure" (2021) in the Annual Review of Ethics, National Institute of Health, https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/case_studies-2021.pdf

Who Is An Author?

A case about Susan Jacobs suddenly finding many co-authors on her paper.

Susan Jacobs, a Ph.D. student sets up, as part of her dissertation, a six-month internship at a prestigious larger institution in order to learn a new molecular technique. Jacobs contacts the laboratory leader, Dr. Marvin Frank, a world-renowned scientist, in the hope of developing new skills for her research and to foster a relationship with Frank who is well connected in her field.

At Frank's laboratory, Jacobs participates in the weekly laboratory meetings, in which everyone participates in a free exchange of ideas about the ongoing projects. In addition, she meets weekly, one on one, with Frank, who provides significant scientific advice and one or two recommendations, which advance her work and move her in a slightly different direction. She discusses the results of her research with her mentor, Dr. Melissa Seabrook, back at her home college, interactions that also push ahead the project she started in Seabrook's lab. Jacobs makes great progress during the six months she spends in Frank's laboratory, and she writes a paper reflecting some important findings. Jacobs puts herself down as first author, Frank as second author, and Seabrook as last author on the paper. At the end of the paper, she gives an acknowledgment to a technician who showed her several techniques and worked with her on a few experiments. Jacobs based her listing of authors on her understanding of the guidelines (aka the Vancouver guidelines) put forth by the International Committee of Medical Journal Editors (ICMJE).

Question 1: Is Jacobs correct to assign authorship herself, or should she and Frank have discussed authorship issues when she started working in his laboratory?

Question 2: What is the difference between an acknowledgment and a listing as an author?

Jacobs gives Frank and Seabrook a draft of her manuscript for review. Frank sends his comments back to Jacobs and changes the authorship listing to include Jacobs, the technician, two postdocs in his lab, two graduate students in the lab, himself, and Seabrook. Frank also gives a copy of the draft to all the members of his laboratory for discussion at the next meeting.

Jacobs is shocked that Frank added the other laboratory members to the draft, explaining to him the Vancouver guidelines and maintaining that the major intellectual and physical work in preparing the paper was done by her and by Seabrook and Frank. Dr. Frank is equally surprised by Jacobs's opinion, responding that he and Jacobs benefited from the input of all the other lab members. Frank adds that a graduate student in the laboratory, Lisa Bain, is writing a short paper that is based on some very exciting preliminary findings, and that Jacobs would be included in the list of authors. Frank says that the results of Lisa's research would need further elaboration in the laboratory and that a second paper using the same data and additional studies would be more comprehensive, and that Jacobs would be included on the second one, too.

Question 3: Although many journals subscribe to the Vancouver guidelines, many do not, and many researchers do not follow the practices that it recommends. Do researchers in your department follow the Vancouver guidelines? If not: what are the most common practices?

Frank insists to Jacobs that the contributions of all the laboratory members were sufficient to satisfy the Vancouver guidelines for both papers, adding that the idea of a scientist acting as an independent entity is an outdated concept and that those who work around a scientist contribute significantly, helping him or her to function. Jacobs tells Frank that she does not want to be included

on Lisa's paper, feeling that she did not contribute adequately. Seabrook, who is in line with Jacobs's understanding of the Vancouver guidelines but was intimidated by Frank's stature, advises Jacobs not to rock the boat, to use Frank's revisions and some of the changes suggested during the laboratory review and to submit the paper to the journal with the authorship he suggested.

Question 4: Who among the authors takes responsibility for submitting the paper to a journal and following up with the editor and peer-review revisions?

Question 5: What are some potential problems with Frank's submitting a paper on preliminary findings and not performing sufficient corroboratory experiments?

Question 6: What kind of problems may arise if the same data is used in multiple papers in the research literature?

Question 7: What might happen if someone is listed as an author on a paper for which he or she did not do any work?

Question 8: What might have been done to resolve Ms. Jacobs's ethical dilemma with Dr. Frank about the authors on the paper?

Credits: https://ccnmtl.columbia.edu/projects/rcr/rcr_authorship/case/index.html#1, which acknowledges the following source: "When in Rome: Conventions in Assignment of Authorship", *Research Ethics: Cases and Commentaries*, Volume 2, Section 1, Authorship., Brian Schrag, ed. Association for Practical and Professional Ethics, Bloomington, Indiana, February, 1998, Prepared under NSF grant No. SBR 9241897