The Stapel Fraud in the Light of Philosophy of Science[1]

Ahti-Veikko Pietarinen

ahti-veikko.pietarinen@helsinki.fi

Professor of Semiotics, Department of Philosophy, History, Culture and Art Studies

University of Helsinki, Finland

Professor, Chair of Philosophy, Ragnar Nurkse School of Innovation and Governance

Tallinn University of Technology, Estonia

Research ethics and philosophy of science meet in one of the most disturbing and wide-ranging scientific frauds to date, the Stapel Case of the Netherlands. I will argue that, beyond the obviously blatant violation of research ethics concerning data cooking and deceit, the nature of the case raises issues about the implications of some relatively widespread issues in the philosophy of science. In particular, I want to point my finger at Kuhnian philosophy of science, or rather its later and largely derailed interpretations that came to be allied with the strong version of the underdetermination thesis. But let me briefly describe the facts of the affair first.

Diederik Stapel was a social psychologist who had worked at Tilburg University in the Netherlands as professor, dean, and principal investigator of large research projects. Previously, he had been employed by the University of Groningen and received the doctorate from the University of Amsterdam in 1997. The topics of his research covered numerous areas in experimental psychology and behavioural sciences, such as social interaction, perception and self-image, as well as consumer behavior, marketing and advertising. Since the late 1990s, he had acquired considerable research funding and support, supervised 20 doctoral theses and published over 150 papers. Many of his publications were published in well-respected journals in the area of social psychology and beyond, including the news-breaking paper in *Science* on how messy environments – including your own office – give rise to discrimination, or one according to which meat-eating was behind human selfishness

[1] Research supported by the Estonian Targeted Funding Scheme SF 014006, *Challenges to State Modernization in 21st Century Europe*.

and aggression, or that we use better manners when a wineglass is on the table. Yet another one managed to conclude that beautiful people are more successful in life.[2]

Praised by his colleagues for his deep knowledge of theories and capacity for innovation and hypothesis generation, it came as a surprise to many when it was reported in 2011 that Stapel had in fact fabricated data for most of his experiments. With little suspicion by the colleagues, editors or referees of his publications, he was able to continue flawed research until finally expelled by Tilburg University in late 2011. The vegetarian joy did not last for long after all. Of course, there are wider implications beyond individual careers, such as the corruption of the scientific literature on the relevant fields and the reputation of those who relied on it, including 14 tainted PhD theses. The self-cleansing mechanisms of science will be shrieking here.

In November 2012, three investigative committees appointed by these universities concluded what was been one of the most thorough investigations into scientific fraud to date. The joint committee report established that Stapel had cooked up the experimental data for at least 55 of his scientific publications (LNDC 2012), now in the process of being retracted. But the alarming fact here is not so much the scale and duration in which such flawed practices took place but their nature. The joint report stages a typical pattern in Stapel's research. First, he would observe some phenomenon concerning the relevant field of investigation. Then, a research question was formulated in the light of the chosen theory. Third, a hypothesis to account for the phenomenon in question was formulated, either by Stapel alone or in collaboration with his students. Following this, an experimental set-up to test the hypothesis in question was also described, and even the rewards to be delivered to the subjects were acquired.

So far so good, as this process indeed resembles the way scientific research routinely is carried out. However, following these steps a completely fictional phase would follow, during which Stapel would invent or massage entire datasets. Experiments and data collection claimed earlier in the course of investigation hardly took place at all. Stapel would just sit and type that data in in private. Even the rewards would not be handed out to the alleged subjects but were just dumped into containers.

[2] I refrain from citing these flawed papers.

The report also established that the conduct continued throughout much of Stapel’s career. Data manipulation in some form or another was likely to have begun already during the course of his PhD studies. But the deceit became increasingly widespread and foolhardy, until the repeated suspicions harboured by his students and colleagues were voiced to the university administration and finally heard and acted upon, and the thorough investigations now completed were launched. As it requires certain skill and effort to invent numerical data in order to support particular hypotheses and not to be caught outright, Stapel would also systematically refuse to disclose the raw data for further inspection and scrutiny, often by lies, threat and intimidation.

Given the systematic nature of data cooking, it lies beyond doubt that Stapel’s case would be an instance of mere faulty judgment or human error. It certainly also is not the case of what one might be taken to pertain to a strategic or interpretive (and not necessarily unethical) choice of, say, omitting data, a practice which may take place when certain types of data may not be appropriate or representative enough in the light of particular research questions at hand and could thus be discarded as it would be clear that they would not establish a particular hypothesis, as might have been the case in the infamous Millikan oil-drop experiment (Hintikka 2005).

So what on earth was this guy thinking about? I am not very interested in this psychological question. Instead, it is more pertinent to ask what factors beyond the mere inborn criminality of mind might have driven an individual to seek a shortcut to academic stardom through spinning, fudging and outright fraud.

What I want to argue here is that the overall increase in fraudulent practices that we have witnessed of late have become a symptom of two largely questionable and in many respects faulty ideas in philosophy of science having gone into their extremes: Kuhn’s or his posse’s willingness to displace facts and evidence with subjective interests and points of views, and Quine’s thesis of underdetermination of theory by evidence under its strong interpretation. Stapel’s case fits in with the pattern of strong underdetermination, and no more blatant and direct example of subjective interests driving the inquiry can be found. According to his own assertion, “the freedom we have in the design of our experiments is so enormous that when an experiment does not give us what we are looking for, we blame the experiment, not our theory. (At least, that is the way I work). Is this problematic? No” (Stapel 2000, quoted in

LNDC 2012, p. 40). What goes under such freedom here? This guy in fact had directly admitted early on that he has no qualms accepting the strong thesis of underdetermination.

That subjective and social interests influence the results does not imply, perhaps with the exception of the version of sociology of science that takes determination of scientific beliefs to be the result of social causes (see e.g. Bloor 1991), that *all* interest-driven scientific practices would be irrational or fraudulent. Kuhn himself was quite clear, especially in his late 1993 "Afterword", that it is an error to suppose that interest-laden research is nowhere compatible with rationality (Kuhn 1993, p. 339).

These two proposals, by Kuhn and Quine, and largely taken in the senses that derive from their later (mis)interpretations, nonetheless sparked off much of the subsequent industry in social-constructionist epistemology and contemporary sociology of science and science studies. These fields are dominated by relativist accounts concerning the nature of scientific beliefs. Kuhn's philosophy of science was not overtly relativist, however. Yet it was extremely limited in the way it posed the questions about the nature of science. Kuhn's interest was in describing, in a characteristically loose way, what kinds of historical developments have taken place in science and how they may have contributed to the theory formation. He was exercising history and sociology rather than philosophy of science. Kuhn did not analyse the methodologically crucial concepts needed to understand the nature of scientific discovery.

The strong sociological programme then took a step away from Kuhn. But I do not take it a plausible account of science either. The sociological programme falters on the inadequacy of the symmetry principle. That principle states that the style of explanation in the sciences must be the same for the true and the false, as well as for the rational and irrational (or non-rational) beliefs (Bloor 1991). I take this to be mistaken. In sciences, and especially in much of the contemporary empirical research resorting to statistical methods, the style of explanation is in fact very different depending on whether the hypotheses entertained were to be true as opposed to them being false. I will return to this point shortly.

Another weak spot of the Kuhnian programme is its pontification on the importance of history of science to the philosophy of science. History is certainly important, but emphasizing this fact was neither particularly significant nor novel. Kuhn was mainly posing the question in response to the positivists' avant-garde break-off with the tradition. Choosing your enemies wisely – and not necessarily the real enemies – is of course the path to success also in academia. Kuhn then rightly notes that the aim of science is not to collect a body of facts constitutive of scientific knowledge and that the practice of actual scientific work and the behavior of the community of inquirers is to be looked into. But that observation was nothing new as it was made long before Kuhn by the American pragmatists, most notably Charles Peirce, for whom, just as for any other serious scientist aiming at an understanding of the content of scientific assertions through fixing the beliefs by eradicating doubt, inquiry does not concern infallible and eternal knowledge (Peirce 1877).

What Kuhn would in his "Afterword" submit to is the denial of the meaningfulness of simultaneously holding the two claims: first, that "successive scientific beliefs become more and more probable or better and better approximations to the truth", and second, "to suggest that the subject of truth claims cannot be a relation between beliefs and a putatively mind-independent or 'external' world" (Kuhn 1993, p. 330). The true impact of this insightful submission has invariably been neglected by the soi-disant Kuhnians. But the idea behind it was already formulated, and in a much deeper sense, by Peirce, who well realized that novel scientific hypotheses can hardly be assigned probabilities at all and that it is thus misleading to think of scientific beliefs as increasingly better approximations of truth. Genuine scientific hypotheses are the ones that are prone to very sudden rebuttals no matter how widely and how long they have been upheld by the community, if they are false.

Now Peirce was a realist in the sense of the denial of the latter clause. That is, realism for him consists in the mind-independence of truth claims. This amounts to an extremely appealing position. What you can simultaneously hold are the negations of these two clauses that Kuhn presents in the "Afterword": you can rest content with being a realist as per mind-independence of scientific assertions, *provided that you deny* the view that scientific beliefs acquire increasingly higher and more accurate probabilistic assignments.[3]

Another idea in which social factors have been proposed to enter the formation of scientific beliefs is the underdetermination thesis. That the available observational evidence may not decide between rival hypotheses[4] is well-known and gives rise to weak underdetermination.

[3] This does not contradict Peirce's limit view of truth as the final opinion that would be agreed upon by the scientific community, since there is no assumption that it is through better and better probabilistic approximations that such views ought to be reached.

[4] Hypotheses are often confused with theories in the literature on philosophy of science but they are not at all the same thing.

More interesting and worrying here nonetheless is the strong thesis of underdetermination, according to which any hypothesis has an incompatible rival to which it is empirically equivalent, that is, makes equal empirical predictions. How does this relate to scientific fraud and data cooking? Well, for no one subscribing to the strong thesis would it be terribly sensible to sweat over seeking for the data to confirm or refute the hypothesis, as one would not be seeking for an understanding of theories at all, merely their possibility or perhaps some predictive power. And for some it indeed does not seem to be a giant leap to be taken from such a position to an altogether corrupt scientific conduct.

One response to the strong thesis has been to recognize that it is not in the nature of science to take observables to be theory neutral. This response in on the right track but it needs to be understood in a specific way, however. Science works with a set of background assumptions. Physics works with systems, not the world at large, and defines the boundary conditions for them which determine the range of applicability of physical concepts. Boundary conditions identify the systems and make them behave independently from the surroundings. This makes an ingenious design of experiments possible. Similar background assumptions permeate other sciences, too.

Thus even in a somewhat stronger sense, it is not possible to argue for an empirical equivalence of hypotheses unless there would be an algorithm that could show that, for any set of background assumptions or boundary conditions, the predictions yielded would in fact be exactly the same. But the existence of such algorithms is precisely what the strong sociologist programme is attempting to deny. Here again, empiricists and sociologists of science are in fact largely occupying the same ground. But hypothesis selection involves a lot more criteria, qualities and parameters which go well beyond the merely empirical or even the social, such as the economy of research, including simplicity (that is, test the cheapest hypotheses first because most hypotheses are likely to turn out false anyway, which is not the same as the simplistic Occam’s Razor), the breadth or coverage of hypotheses, and the fruitfulness of hypotheses to suggest further ideas, given that most of our hypotheses will in any case turn out false. Theory development involves peculiar forms of reasoning where previous laws come to be merged and where the empirical and mathematical contributions are inherently intertwined. Go nowhere further than to Einstein’s discovery of the famous formula which he reasoned out from the laws of the conservation of energy and the conservation of mass. That equation has endless applications to what physicists call systems, those independent parts of the world that define the range of applicability. Thus formal differences in the meaning of empirical predictions are so important. Thus not just observations but also experiments are theory-laden. And thus it is meaningless to talk about empirical equivalence independent of the theory. This, in a nutshell, is my argument showing the vacuity of the strong thesis.

A further problem in typical explanations arrayed by the sociology of science is the assumption that the reference points that empirical scientists posit and which they do need in order to vouchsafe objectivity and comparability of scientific assertions, are infallibly true. However, this is to simplify the picture of scientific practice to the point of invalidity. The reference points typically have some independent evidence and reason for their existence which is not tied with particular hypotheses an empirical scientist attempts to confirm.

*   *   *

Needless to say, teaching philosophy of science is vitally important to achieve the goals of general science education. But unfortunately, philosophy of science has been contaminated by movements that do not strive to the understanding of the real content of scientific work.[5] Courses in philosophy of science are often taught by scholars who do not endeavour to explain the nature of scientific practice, or the methodological tools employed, or the semantics of the key terms involved in the investigation. I feel that philosophers, sociologists or historians of sciences may in fact not the best persons to achieve that knowledge. Kuhn and Quine were outsiders to real science. They did not aim at really understanding it. The great science popularizes, among them names such as Darwin, Ortega y Gasset, T.H. Huxley, Asimov, Einstein, Feynman or Russell, did a much better job in educating others about philosophically central issues in and around science, society and culture. They can also tell interesting stories about the history of discovery in science and technology. The greatest philosophy *and* history of science is often made by the philosophically minded scientists or scientifically minded philosophers and writers, witness for instance the essays in the marvelous collection of Gardner (1994). To understand the relevance of history to science, listen to those who made history.

But it is neither the resentment nor envy of science that keeps an honest inquirer busy at work. It is the potentiality of hypotheses turning out false, however inconvenient or painful

[5] Pietarinen (2012) presents pragmatist arguments against social constructionism in the area of organizational semiotics.

that may be. Losing sight of the truth guiding the inquiry may deliver wrong causes unworthy of fighting for. Though truth in the limit is something never actually reached, its dismissal is unsafe.

Last, let me bring up an acute issue concerning the nature of statistical research in contemporary empirical sciences. I mentioned this in the beginning as telling against the symmetry principle assumed in the strong programme of the sociology of science, that is, that the same style of explanation should be employed for the true as well as for the false hypotheses. Now in applied statistics, it is not too difficult to find dependencies from almost any dataset. However, only very few of these dependencies are true. By the very nature of the way the world is, most of them will be false. But in our media-driven day and age, the dependencies, whichever way they actually are, tend to be taken newsworthy to be published as showing something significant about the results of the investigation. In bookstores, you do not find popular economics books entitled “How I Failed to Make My First Million” or “Anybody Can Be a Loser”, though such accounts would be much more helpful in practice. If a false dependency fails to be disproven, increasingly many researchers seem to feel the temptation to publish that fact as showing something interesting about the case, something which becomes marketed as a dependency (‘eating cereals improves cardiovascular health’) but which in fact is a false positive. The positive bias is increasingly widespread in contemporary psychology and the social sciences playing with large datasets.[6]

Yet science, which in these days is indeed largely based on statistical analyses, tells strongly against positive bias. After all, the best hypotheses in science are those that not only suggest themselves for the investigation, but are at the same time the most prone to be very suddenly rebutted, *if* they are false. Sooner or later, a scientist should be able to catch and debunk them, provided that they are false. If this is what Kuhn actually meant by his superficial story

[6] Fanelli (2012) investigated 4600 papers across disciplines and established that the percentage of papers reporting a positive support for a particular tested hypothesis had increased over 22% on average from 1990 to 2007, the increase being highest in the social and biomedical disciplines. It is not a plausible explanation to this trend that the statistical methods themselves in these disciplines would have improved so much as to match the corresponding ability of contemporary research to catch the true positive dependencies increasingly well. On the contrary, the explanation for the tested hypotheses to be increasingly likely to be true suggests that increasingly more ‘safe’ hypotheses are selected to guarantee not only publication but also funding and publicity. Fanelli (2010) reports that in psychology, psychiatry, economics and business the chance of reporting false positives is five-fold compared to space science, 3.4 times higher in behavioural and social sciences in comparison to physical and chemical sciences, and 2.3 times higher in social sciences compared to physical sciences.

involving sudden changes in paradigms, then the point is already there in Peirce's 19^{th} century logic of science.

Unfortunately, this time-honoured bedrock of science seems to have become increasingly forgotten due to a multiplicity of factors. Among them are the intolerably high level of competition in contemporary academia, harsh socio-economic conditions to pursue truly long-term goals, salami projects, eerie systems of academic advancement and certification, lax peer review system, exhaustion of creativity and imagination, disdain for critical thinking and negative results, or even such methodological issues such as the widespread yet largely unjustified uses of Bayesian update methods to measure degrees of beliefs in epistemology or in economics. Who would fund a fifteen-year project that would show, once and for all, that sociological approach has no bearing on philosophy of science, that psychology is irrelevant to the nature of reasoning, or that feminism has no bearing on epistemology? (It would not take that long, though.) Susan Haack (1997) talks about anti-science in the age of preposterism. Sham and fake reasoning is surely all around us. And, as I wanted to emphasise here, we need to include here the escalation of unethical conduct which in part may have been due to an uncritical acceptance of forms of radicalized philosophy of science that neglects the worry about the sham and the fake and takes good reasoning only as an optional ornament.

In the light of the scams of the magnitude of the Stapel Case, I here raise the issue of whether the relativist ideas – which originally might have looked as much more innocent playgrounds for those who did not suspect the worst – have now shown their grotesque sides in the proliferation of pseudo-inquiry.[7] It may not be too surprising after all that Quine would prefer behaviourist psychology over the tools of reasoning, logic and semantics, if the underdetermination thesis is to hold any water. Kuhn arrays some vague historical and sociological concepts instead of even trying to make their meanings precise and useful to scientists and scholars themselves.

[7] And recall that, despite the incommensurability idea and the fact that he did not define realism as mind-independence or the nature of reality as a language-independent entity, Kuhn himself was reluctant to characterise himself as an anti-realist. It was his soi-disant followers that used Kuhnian interpretations for relativistic causes. Equally, although he may not have had a calculated intention to blur the distinction between science and non-science, the 'Kuhnians' may have accomplished just that.

These remarks suggest that the sad case of Stapel can be classified among the anti-scientific pontifications that include those "post-modern literary intellectuals" arrayed and knocked off in Sokal (2008). What the literary intellectuals have achieved is the public reception of the obscure as the profound, or the fashionable as the noteworthy, which is enough of a motivation for the sham and the fake scholar to advance herself and not the truth.

References

Bloor, David (1991). *Knowledge and Social Imagery*. Chicago: Chicago University Press.

Fanelli, Daniele (2010). "'Positive' Results Increase down the Hierarchy of the Sciences", *PLoS ONE* 5, e10068. doi:10.1371/journal.pone.0010068.

Fanelli, Daniele (2012). "Negative Results are Disappearing from most Disciplines and Countries", *Scientometrics* 90, 891-904.

Gardner, Martin (1994). *Great Essays in Science*. New York: Prometheus Books.

Haack, Susan (1997). "Science, Scientism, and Anti-Science in the Age of Preposterism", *Skeptical Inquirer* 21.6.

Hintikka, Jaakko (2005). "Omitting Data: Ethical or Strategic Problem?", *Synthese* 145, 169-176.

The Levelt, Noorth and Drenth Committees (LNDC) (2012). "Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel", 28 November 2012. https://www.commissielevelt.nl/

Kuhn, Thomas (1993). "Afterwords", in *World Changes: Thomas Kuhn and the Nature of Science*. Paul Horwich (ed.), Cambridge: Mass.: MIT Press.

Peirce, Charles (1877). "Fixation of Beliefs", *Popular Science Monthly*.

Pietarinen, Ahti-Veikko (2012). "On the Conceptual Underpinnings of Organisational Semiotics from the Pragmatist Point of View", *Information Systems Journal*, in press.

Sokal, Alan (2008). *Beyond the Hoax: Science, Philosophy and Culture*. Oxford: Oxford University Press.

Stapel, D.A. (2000). “Moving from Fads and Fashions to Integration: Illustrations from Knowledge Accessibility Research”, *European Bulletin of Social Psychology* 12, 4-27.