Published in final edited form as:

Sci Eng Ethics. 2014 June ; 20(2): 423–431. doi:10.1007/s11948-013-9466-z.

Data Fabrication and Falsification and Empiricist Philosophy of Science

David B. Resnik, JD, PhD

National Institute for Environmental Health Science, National Institutes of Health, 111 Alexander Drive, Box 12233, Mail Drop CU03, Research Triangle Park, NC, 27709, USA,
resnikd@niehs.nih.gov Phone: 919 541 5658 Fax: 919 541 9854

Abstract

Scientists have rules pertaining to data fabrication and falsification that are enforced with significant punishments, such as loss of funding, termination of employment, or imprisonment. These rules pertain to data that describe observable and unobservable entities. In this commentary I argue that scientists would not adopt rules that impose harsh penalties on researchers for data fabrication or falsification unless they believed that an aim of scientific research is to develop true theories and hypotheses about entities that exist, including unobservable ones. This argument presents a challenge for constructive empiricists, such as van Fraassen. Constructive empiricists need to be able to explain why rules pertaining to data fabrication and falsification do not threaten their philosophy of science.

Keywords

Data fabrication and falsification; empiricism; realism; observation/theory distinction; ethics; scientific practice

The philosophical debate between realists, such as Boyd (1983) and Chakravartty (2007), and constructive empiricists, such as van Fraassen (1980, 1985, 2001), focuses on the attitude one should take toward theories and theoretical entities. Realists hold that the aim of science is provide us with a true description of the world and that to accept a scientific theory (or hypothesis) is to believe that it is true and that the entities described by the theory, including entities we cannot observe, exist. Constructive empiricists hold that the aim of science is to provide us with an empirically adequate description of the world and that to accept a scientific theory (or hypothesis) one need only believe that it is empirically adequate: one need not believe that unobservable entities described by the theory exist. Both realists and constructive empiricists agree that there is a mind-independent external world and that physical objects exist; they are not idealists or relativists. They differ in how they conceive of the extent of our knowledge beyond the observable realm and our metaphysical commitments to unobservable entities (Chakravartty, 2007).

Most scientists pay scant attention to metaphysical issues, such as the realism/empiricism debate (Fine, 1996), but are significantly concerned about ethical ones, such as fraud, plagiarism, authorship attribution, conflict of interest, and protection of human and animal subjects in research (citation omitted for review). Although fraud is thought to be rare in

science (Fanelli, 2009), it draws considerable scrutiny from researchers, institutions, and sponsors because it can have far reaching negative consequences. Fraudulent research misleads scientists by sending them down blind alleys, destroys trust among researchers, sabotages scientific collaborations, weakens the public's support for research, and can cause considerable harm to society. For example, faked research data may lead to the approval of unsafe drugs or the construction of dangerous buildings.

Scientists who are caught conducting fraudulent research may face adverse social, financial, and legal consequences. The U.S. government prohibits misconduct in federally-funded research, which is defined as “fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results (Office of Science and Technology Policy 2000, p. 76262).” Scientists who are found to have committed any of these dishonest acts can face an array of penalties, including termination of employment by their institutions and a ban on receiving funds from government agencies. They may also face criminal penalties if they are convicted of defrauding the government (citation omitted for review). Most scientific journals have policies that prohibit fraud and require authors to retract or correct papers associated with an official finding of misconduct (citation omitted for review). Professional associations have also developed ethics codes that strongly endorse honesty in science. Is there any relationship between concerns about fraud in science and metaphysical issues, such as the realism/empiricism debate? In this commentary, I will explore the relationship between rules pertaining to fraud in research and debates about scientific realism. What are the philosophical commitments involved in taking an ethical or legal stance against fraud in science and what implications, if any, does that stance have for the dispute between realists and empiricists?

A distinction between observation and theory is a key pillar in constructive empiricism's philosophy of science, since constructive empiricists argue that we cannot know whether theoretical entities (i.e. things we cannot observe with the unaided senses) exist. Constructive empiricists do not claim that theoretical entities do not exist. Instead, they take an agnostic attitude toward the existence of atoms, electrons, deoxyribonucleic acid (DNA), cells, and so on (van Fraassen 1980, 1985).

Assuming empiricists can draw a coherent observation/theory distinction, how should they respond to rules concerning data fabrication and falsification? Does claiming that data have been fabricated or falsified commit one to believing in the existence of entities described by the data? Before we can address these questions, it is necessary to explore the concept of data in greater depth. While philosophers have paid considerable attention to the concept of observation, scientists rarely use the word ‘observation’ when reporting results but talk instead about data (Bogen and Woodward, 1988). Data, the backbone of scientific inference, can be defined as recorded information used to support hypotheses, theories, or models in science. Data may be generated when human beings record their observations or when machines produce outputs. For example, if a zoologist observes a rare primate species in the wild, her recorded observations would be data. Data could also be produced by an automated deoxyribonucleic acid (DNA) sequencing machine that analyzes a biological sample.

Data often undergo several stages of processing before they are presented in scientific papers or reports. For example, consider functional magnetic resonance imaging (fMRI) data of brain activity used in neuroimaging research. The images reported in scientific papers provide an anatomical picture of the brain in black, gray, and white. Different colors, such as yellow, red, and blue, in the image indicate levels of metabolic activity in different areas of the brain. To produce these images, powerful magnetic fields are applied to an individual's brain. The magnetic fields cause protons in hemoglobin molecules in the brain to emit radio signals. When the magnetic field weakens, the radio signals from highly oxygenated areas of the brain (with more hemoglobin) deteriorate at a slower rate than radio signals in areas with less oxygen. Computer programs analyze the radio signal data to produce images with different colors corresponding to different levels of oxygenation or brain activity (Bogen, 2009).

Data are often highly theory-dependent. First, scientists often use theories to generate data. For example, theories concerning magnetic fields, radio emissions, atomic and subatomic physics, and cellular metabolism are essential to producing fMRI images. Second, data often report information about things we cannot directly observe. In biomedicine, data may pertain to DNA, ribonucleic acid (RNA), proteins, and other macromolecules; cellular processes, such as cell signaling, cell death, and cell division; inflammatory responses; oxidative stress; and tissue damage.

Thus, the relationship between scientific data and observation is not straightforward (Bogen and Woodward, 1988). A simplistic way of construing this relationship would be to say that data and observations are one and the same. Presenting data to support a theory or hypothesis is the same as reporting observations to support the theory or hypothesis. While realists may have no qualms about this way of viewing the relationship between data and observations, empiricists should take issue with it, because empiricists rely on a distinction between observation and theory and are skeptical about the existence of theoretical entities. Empiricists should not regard DNA sequence data, fMRI data, or other types of theory-dependent data as observations, unless they are willing to expand the scope of what can be observed. At the very least, empiricists must provide an account of the relationship between data and theory, and whether any types of scientific data should be treated as observations.

With this account of the relationship between observation and data in mind, we can now consider the philosophical import of rules pertaining to data fabrication and falsification. Data fabrication and falsifying both are forms of lying about the data reported in a scientific paper (citation omitted for review). Lying involves making a statement intended to mislead others, which could include making a statement that one knows or believes to be false or making a statement that omits some important information (i.e. not telling the whole truth) (Bok 1979). The U.S. government defines data fabrication as “making up data or results and recording or reporting them” and data falsification as “manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record (Office of Science and Technology Policy 2000, p. 76262).” For example, a scientist who claims in a paper that he tested a chemical on 100 rodents but only used 50 and made up the data for the other 50 would be fabricating

data. A scientist who tests a chemical on 100 rodents but omits or changes results from 50 rodents, to provide better support for his hypothesis, would be falsifying data.

Most scientific papers include a materials and methods section that explains how data were produced/acquired and analyzed. It is important to carefully describe how data were generated/acquired and analyzed in a paper or report so that other scientists can validate the work and replicate experiments (citation omitted for review). Data fabrication and falsification involve not only lying about the data, but also lying about how the data were generated/acquired or analyzed.

One of the most infamous cases of data fabrication in recent memory occurred when Seoul National University scientist Woo Suk Hwang and his research team published two papers in *Science* (Hwang et al, 2004, 2005), claiming to have produced patient-specific embryonic stem cells through somatic-cell nuclear transfer (SCNT). In SCNT, the nucleus is removed from an unfertilized egg and a nucleus from a somatic cell is inserted into the egg. The donor nucleus reprograms the egg, which can begin cell division. Stem cells can be harvested from the developing embryo to create cell lines for transplantation. If the embryo is implanted in a womb, the resulting offspring would be a clone of the individual that donated the nucleus. SCNT has been performed in animals successfully to produce cloned sheep, dogs, and mice, but it has not been performed in humans. Hwang's research, if substantiated, would have represented an important advance in the field of regenerative medicine, since cell lines produced by this process would be immunologically compatible with the patient's body, which would reduce the risk of tissue rejection (citation omitted for review).

Though Hwang was hailed as a national hero in South Korea following the publication of the two papers, suspicions concerning the legitimacy of his work emerged in the fall of 2005, when an anonymous informant told the South Korean investigative news program PD Notebook that his research was fraudulent. PD Notebook began investigating the case, and Seoul National University launched its own investigation. A university committee concluded in December 2005 that Hwang's data were faked. Hwang's papers included microscopic images of human cell lines that he said were produced by SCNT. Hwang also provided data concerning other characteristics of the cell lines, including genetic and immunological analyses. The committee asked three laboratories to perform tests to determine whether the cells reported in the paper matched the cells from the donors. The laboratories compared DNA from cell lines reported in the papers to DNA from the donors and found that they did not match. Hwang was dismissed from Seoul University, barred from receiving research funding from South Korea, and convicted of embezzlement and bioethics violations. He was sentenced to two years in prison, though his sentence was suspended. *Science* retracted both papers. The incident caused considerable embarrassment for South Korean researchers and had a negative impact on the public's perception of stem cell research (Cyranoski, 2006, citation omitted for review).

Hwang's papers reported fabricated data pertaining to things that are not observable with the unaided senses, including cells lines, blastocysts, embryos, DNA, and histocompatibility complexes (proteins on the surface of cells) (Hwang, 2004, 2005). He received a harsh

punishment for his fabrications. Hwang is not the only person to have received a prison sentence for scientific fraud, however. In 2006, University of Vermont researcher Eric Poehlman was sentenced to serve a year and a day in federal prison for defrauding the federal government as a result of fabricating or falsifying data on fifteen federal grant applications worth \$2.9 million and seventeen publications over a ten-year period (citation omitted for review). Although most researchers do not go to jail for fabricating or falsifying data, they may face other career-threatening consequences, such as loss of employment or a ban on receipt of federal funding.

The harsh sanctions imposed on researchers who are found to have fabricated or falsified data pertaining to unobservable things present a potential problem for empiricism. Hwang lost his job and was sentenced to prison for faking data pertaining to stem cells. This seems like an onerous penalty to impose on someone if one does not believe that one of the goals of stem cell research is to determine the truth about stem cells. The fact that scientist impose tough sanctions on colleagues who lie about data that describe unobservable things supports the view that scientists have realist aims. They are interested not only in obtaining truth about the observable realm (i.e. empirical adequacy) but also in obtaining truth about things we cannot directly observe, such as stem cells.

Suppose that Hwang were not a scientist working for the government but were instead a science fiction writer who had published a novel on stem cell research. If this were the situation, he would not be charged with data fabrication, even if his novel included pages of fictional data. People who write works of fiction do not face charges of data fabrication or falsification because the audience does not expect a work of fiction to report the truth. When people read a work of fiction, they understand that statements contained in the work are not intended to be true and that the people, places, and things described in the work may not exist. This is not the case when scientists, funders, and others read scientific papers and reports. Fiction does not deal with reality, but science does.

This problem for empiricism can be stated more formally as an argument:

1. Scientists have rules pertaining to data fabrication and falsification and these rules are enforced with significant punishments, such as loss of funding, termination of employment, etc.
2. Scientific data frequently describe entities that are not observable with the unaided senses.
3. If scientists have rules pertaining to data fabrication and falsification that are enforced with significant punishments and the data frequently describe unobservable entities, then they believe that an aim of scientific research is to develop true theories and hypotheses about entities that exist, including unobservable ones.
4. Therefore, scientists believe that an aim of scientific research is to develop true theories and hypotheses about entities that exist, including unobservable ones (from 1, 2 and 3).

5. We should not adopt a philosophy of science that is contrary to practice of science (i.e. what scientists believe and do) without good reason.
6. Constructive empiricism holds that the aim of science is not to develop true theories and hypotheses concerning unobservable entities but to develop theories and hypotheses that are empirically adequate (van Fraassen's definition of empiricism).
7. Constructive empiricism is contrary to the practice of science (from 4 and 6).
8. Therefore, we should not accept constructive empiricism (from 5 and 7).

Constructive empiricists have several ways of responding to this argument. First, they can deny the third premise by claiming that the rules concerning data fabrication and falsification pertain to what is observable; penalties are imposed on scientists for making claims about things that can be observed. There is no need to assume that an aim of science is to develop true theories or hypotheses concerning unobservable thing since data fabrication and falsification have an adverse impact on science's claims about what we can observe. Thus, scientists can condemn data fabrication and falsification without assuming that science has realist aims.

This is a perfectly reasonable response to science's rules pertaining to data fabrication and falsification, if one assumes that data only describe observable entities. An empiricist who is willing to expand the range of what counts as observable can legitimately claim that penalizing scientists for data fabrication or falsification involves no commitment to realist aims. However, this does not seem to be the position that van Fraassen and other empiricists. Van Fraassen's view implies that stem cells are not observable. If one adopts this view, it is difficult to justify the harsh penalties imposed on Hwang for fabricating stem cell data.

Second, constructive empiricists could deny the third premise by arguing that data fabrication and falsification is prohibited in science because it is deceptive, and deception is unethical. A scientist who fabricates or falsifies data is not being honest about how the data were generated or analyzed, because the scientist has not followed the methods and procedures described in the paper or report. One does not need to assume that science has realist aims to regard this type of deception as unethical. A scientist who fabricates or falsifies data is not playing by the rules of science.

While it is correct that data fabrication and falsification involve a deviation from the methods and procedures described in a paper or report, there is more to fabrication and falsification than this. The reason why deviation from methods and procedures described in a paper or report is harshly condemned is that it involves lying about the data, which leads to unreliable or unrepeatable results. Hwang's fabricated data could lead scientists to adopt false hypotheses, theories, and beliefs about stem cells. If a scientist made a deviation from the methods and procedures described in the paper that had no impact on the data or results, then this deviation would not be treated as fabrication or falsification. For example, suppose a sociologist said that he used a particular statistical program, such as Statistical Analysis System (SAS), to analyze the data reported in a paper, when he really used a different, but similar program, such as Statistical Packages for Social Scientists (SPSS), but that this had

no impact on the data analysis. If this problem were discovered, he might be asked to submit a correction to the journal, but he would not be charged with fraud.

Third, constructive empiricists can deny the fifth premise by claiming that it is acceptable to adopt a philosophy of science that is contrary to the practice of science (Wylie, 1986). Philosophy of science is a normative discipline that seeks to evaluate and criticize science. By reflecting on the nature of scientific reasoning, language, and knowledge, philosophers can make suggestions for reforming scientific practice. If the rules adopted by scientists penalize researchers for making false claims about things we cannot observe, then perhaps the rules need to be changed, so that they conform to empiricist principles. For example, rules pertaining to data fabrication and falsification would apply only to data that describes things that can be observed. Misreporting or manipulating data pertaining to unobservable things would be contrary to good scientific practice, but it would not be regarded as data fabrication or falsification. Empiricists have long acknowledged that many researchers subscribe to scientific realism, but they seldom take this fact to constitute a sound argument for realism, since philosophical arguments should appeal to considerations that are independent of scientific practice (Fine, 1996).

This response raises a larger issue concerning the philosophy of science: should philosophy of science be normative or descriptive? Prior to Thomas Kuhn's (1962) landmark book *The Structure of Scientific Revolutions*, philosophy of science was regarded as a normative discipline that focused on the logic of science and the justification of scientific hypotheses, theories, and methods (Kitcher 1995). Kuhn and his followers argued that the history, sociology, psychology, and economics of science have an important bearing on the philosophy of science, and that philosophers can ill afford to ignore scientific practice (Lakatos, 1980, Hull, 1990, Kitcher, 1995). Scientific practice teaches us important lessons about the norms, goals, and traditions of research, and how scientists accept, reject, or revise theories and hypotheses (Kuhn 1962).

Kuhn's ideas have transformed the philosophy of science, and most philosophers today recognize the relevance of the practice of science to the philosophy of science (Gieryn, 1985, Laudan, 1984). Acknowledging that the philosophy of science should be accountable to the practice of science does not imply that philosophy of science must faithfully follow every aspect of scientific practice, as this would undermine critical reflection on the practice of science. Philosophy of science still has an important role to play in criticizing or evaluating scientific practice even if it should be attentive to the history, sociology, psychology, and economics of science (Kitcher, 1995).

Since philosophy of science can and should maintain a critical distance from scientific practice, constructive empiricist challenges to the fifth premise in the argument have considerable merit. However, constructive empiricists still must provide a good reason why we should accept a philosophy of science that seems to be inconsistent with the practice of charging researchers with fabricating or falsifying data concerning unobservable things. Constructive empiricists might argue that their view can be justified on independent grounds. For example, according to the pessimistic induction argument advanced by Laudan (1977, 1981), we should be skeptical about the existence of unobservable entities, since the

history of science contains many examples of entities postulated by theorists, such as phlogiston, the ether, a vital force, and epicycles, which scientists no longer believe exist. According to the underdetermination argument advanced by van Fraassen (1980), the proof for any scientific theory is always tentative, because indefinitely many theories may fit the data (i.e. they are empirically equivalent). Since different empirically equivalent theories may postulate the existence of different theoretical entities, we should be skeptical about the ontological claims made any theory.

It may indeed be the case that constructive empiricism can be established on independent grounds that have nothing to do with scientists' attitudes toward data fabrication and falsification. Assessing empiricism as a philosophy of science is an issue beyond the scope this paper (see Chakravartty, 2007). All I have attempted to show in this commentary is that the rules pertaining to data fabrication and falsification in science present a problem for constructive empiricism because they seem to commit scientists to a realist understanding of their work. Constructive empiricists need show why this aspect of the scientific practice does not constitute a credible objection to their view.

Acknowledgments

Omitted for review.

References

- Bogen, J. [Accessed: 9 November 2012] Theory and observation in science. Stanford Encyclopedia of Philosophy. 2009. <http://plato.stanford.edu/entries/science-theory-observation/#DatPhe>
- Bogen J, Woodward J. Saving the phenomena. *Philosophical Review*. 1988; 97:303–352.
- Bok, S. *Lying: moral choice in public and private life*. New York: Pantheon Books; 1979.
- Boyd RN. On the current status of the issue of scientific realism. *Erkenntnis*. 1983; 19:45–90.
- Chakravartty, A. *A metaphysics for scientific realism: knowing the unobservable*. Cambridge: Cambridge University Press; 2007.
- Cyranoski D. Blow follows blow for stem-cell work. *Nature*. 2006; 439:8. [PubMed: 16397471]
- Doppelt G. The naturalist conception of methodological standards in science: a critique. *Philosophy of Science*. 1990; 57:1–19.
- Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*. 2009; 4:e5738. [PubMed: 19478950]
- Fine, A. *The shaky game: Einstein, realism and the quantum theory*. Chicago: University of Chicago Press; 1996.
- Hull, D. *Science as a process*. Chicago: University of Chicago Press; 1990.
- Hwang WS, Roh SI, Lee BC, Kang SK, Kwon DK, Kim S, Kim SJ, Park SW, Kwon HS, Lee CK, Lee JB, Kim JM, Ahn C, Paek SH, Chang SS, Koo JJ, Yoon HS, Hwang JH, Hwang YY, Park YS, Oh SK, Kim HS, Park JH, Moon SY, Schatten G. Patient-specific embryonic stem cells derived from human SCNT blastocysts. *Science*. 2005; 308:1777–83. [PubMed: 15905366]
- Hwang WS, Ryu YJ, Park JH, Park ES, Lee EG, Koo JM, Jeon HY, Lee BC, Kang SK, Kim SJ, Ahn C, Hwang JH, Park KY, Cibelli JB, Moon SY. Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst. *Science*. 2004; 303:1669–74. [PubMed: 14963337]
- Kitcher, P. *The advancement of science*. New York: Oxford University Press; 1995.
- Kuhn, T. *The structure of scientific revolutions*. Chicago: University of Chicago Press; 1962.
- Lakatos, I. *The methodology of scientific research programmes*. Cambridge: Cambridge University Press; 1980.
- Laudan, L. *Progress and its problems*. Berkeley, CA: University of California Press; 1977.

- Laudan L. A confutation of convergent realism. *Philosophy of Science*. 1981; 48:19–48.
- Laudan, L. *Sciences and values*. Berkeley, CA: University of California Press; 1984.
- Office of Science and Technology Policy. Federal research misconduct policy. *Federal Register*. 2000; 65(235):76260–64.
- Scerri, E. Normative and descriptive philosophy of science and the role of chemistry. In: Baird, D.; McIntyre, E.; Scerri, E., editors. *Philosophy of chemistry: The synthesis of a new discipline*, Boston studies in the philosophy of science. Vol. 242. Dordrecht: Springer; 2005. p. 119-128.
- van Fraassen, BC. *The scientific image*. Oxford: Clarendon Press; 1980.
- van Fraassen, BC. Empiricism in the philosophy of science. In: Churchland, PM.; Hooker, CA., editors. *Images of science*. Chicago: University of Chicago Press; 1985. p. 245-308.
- van Fraassen BC. Constructive empiricism now. *Philosophical Studies*. 2001; 106:151–170.
- Wylie A. Arguments for scientific realism: The ascending spiral. *American Philosophical Quarterly*. 1986; 23:287–98.