

Literature Assignment 0**Due: 5pm, September 29, 2015**

Literature assignments will typically deal with papers describing new hypotheses, thought experiments on how to apply new scientific techniques, and discussions about which technique might be the best fit for a particular task. However, in this assignment we ask you to consider how scientific misconduct has affected human society, scientific community, academic institutions, and you as an individual. These articles are focused on scientific research and publication, but many have close analogs in classes such as this one.

Article:

- Lüscher, T.F. *The codex of science: honesty, precision, and truth--and its violations. European heart journal* (2013) **34**, 1018-1023.
- Roig, M. *Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing*, pp.1-15, 41-44, 51-63. Retrieved from the Office of Research Integrity, National Institutes of Health website: <http://ori.hhs.gov/avoiding-plagiarism-self-plagiarism-and-other-questionable-writing-practices-guide-ethical-writing>
- Alspach, J.G. *A Case of Plagiarism: Lessons for Editors Authors, Reviewers, Readers and Plagiarists. Critical Care Nurse* (2014) 34:5, 12-16.

Read this article and briefly answer the following questions. You may read additional materials, if you wish. If you do, you must cite your sources. *You may not quote verbatim without attribution.* This assignment may be submitted in class, in MI646, or electronically to comp-bio@cs.cmu.edu.

1. Aylmer Fisher accused Gregor Mendel of doctoring his data.
 - a. What was the basis of Fisher's accusation?

The empirical results were too close to the numbers predicted by the theoretical model. One might expect more deviation from the ideal due to experimental error, as well as biological features not captured by the model.

- b. What does Luscher consider to be the key issue in assessing whether Mendel committed fraud?

Intent. If Mendel presented the most accurate data possible, given the means available to him at the time, then he did not commit fraud. If Mendel did drop data points, but believed that he had a scientifically justifiable reason for doing so, then he is not guilty of fraud, although he might be misguided.

2. Luscher points out that while Mendel may or may not have falsified his data, Mendel's conclusions were correct. In what way is scientific misconduct harmful, when the conclusions are incorrect? In what way is scientific misconduct harmful when the conclusions are correct?

Incorrect conclusions: If other scientists design studies that are based on the assumption that the results reported are true, they may waste time and resources on projects that are irrelevant or doomed to fail. Doctors may prescribe treatments based on the falsified. These treatments could harm rather than help the patient.

Correct conclusions: Even when the results are correct, scientific misconduct pollutes the scientific process. Junior scientists may conclude that falsification or fabrication is acceptable. Misconduct can lead to unrealistic expectations of scientific output. Scientists who adhere to appropriate conduct may have difficulty competing for funding or promotion.

3. Is dropping data points ever justified? If yes, give an example. If no, explain why not.

Yes. Excluding data can be justified when there is a reason to believe that certain data are not valid (e.g., known contamination, mislabeling, etc.) There are also statistical bases for excluding data; for example, a measurement that is many standard deviations from the mean is considered an outlier and may be excluded.

4. What responsibilities does a scientist incur when s/he agrees to be an author on a manuscript that is submitted for publication?

By placing his or her name on a publication, an author is implicitly taking responsibility for the accuracy and integrity of the scientific research reported in the manuscript.

5. Alspach states that according to the COPE guidelines, it is the duty of the "culpable editor" to ask that a suspected case of misconduct be investigated if the author does not provide an explanation that the editor considers to be adequate. What governing bodies may carry out this investigation? Why might those institutions be expected to carry out a fair investigation? Why might those institutions be unlikely to carry out a fair investigation?

The institution where author of the suspected misconduct is employed or an agency that regulates the industry in which that author works.

Fairness: An institution such as a university has an interest in protecting employees from arbitrary accusations. A university also wants to safeguard its reputation for scholarly integrity. In the best case, the balance between these two interests will result in a fair investigation.

Unfairness: If the investigation is carried out by an institution that does not have an equal relationship to the two authors (i.e., the author suspected of misconduct and the author whose work was allegedly plagiarized), then bias could result. This could arise if the investigating agency is the employer of the suspected author. Investigation by a regulatory agency might be less likely to be inherently biased unless the authors work in an industry where there is a close relationship between the regulators and the regulated.

6. In each of the following scenarios, did scientific misconduct occur? Why or why not? Explain your reasoning.

- a. Professor Victor Bogon of the University of Kansas' Oceanview Campus publishes two separate articles reporting different results. They have different titles and publication dates. There is little similarity between the text in the introduction, results and discussion sections of the two articles. The background sections, however, are identical.

Misconduct occurred. This is self-plagiarism and is unacceptable.

- b. Dr. Bogon uses phylogenetic methods to infer the evolution of North American Squirrels. Dr. Smith's group at the Marshview campus of the University of Arizona has been using similar methods to infer the evolutionary history of the capybara. Each group publishes an article reporting their respective results. The methods sections of the two articles read very similarly, but none of the sentences or phrases are exactly the same. The papers cite many of the same articles. The other sections of the articles are not very similar.

When the language used is very technical, as is often the case in Methods sections, the number of different ways of expressing an idea may be limited. In this case, the two methods sections are probably similar by chance. No misconduct occurred.

- c. One year later, Dr. Bogon turns his phylogenetic gaze to marmots and prairie dogs. In this study, he applies a new method, first described in a publication by Dr. Smith's group, and finds a startling result. Bogon submits a manuscript on this work. In the methods section, Bogon borrows heavily from Smith's capybara paper. Many of the sentences in Bogon's manuscript are identical to sentences in the Smith paper. Bogon cites the Smith article in the introduction to his manuscript.

This is verbatim plagiarism. It is unacceptable to reproduce language verbatim without quotation marks. However, direct quotes should only be used in very specific circumstances, which do not apply here. Bogon should have paraphrased and cited Smith. Further, citing Smith somewhere else in the manuscript is not sufficient. Smith should be cited in the methods section to make it clear that the methods used are based on innovations by Smith.

- d. Dr. Bogon's article about ground squirrels includes a table on a set of "speciation genes" that he hypothesizes contributed to the divergence of Mexican Ground Squirrels and Thirteen-lined Ground Squirrels. In a third manuscript, Dr. Bogon uses this same data in a discussion of the speciation of marmots and ground squirrels. The third manuscript contains a table that includes the data from the original ground squirrel publication and cites that publication in the results section. The table of speciation genes is not specifically associated with a citation to the original paper.

Without a citation, this constitutes double publishing. When submitting the manuscript, Bogon should inform the editors of the journal that this paper contains previously published data as well as new

data. Further, the manuscript should be written in such a way that it is clear to the reader that the data in the table has been previously published.

- e. Dr. Smith's capybara article includes an analysis of predicted genes in the capybara genome. After the manuscript is published, Smith receives an email from Dr. Presley, of Nashville University, pointing out that some of the predicted genes are not valid open reading frames, either lacking start codons or containing premature stop codons. After these incorrectly predicted genes are removed from the analysis, Smith's results are no longer statistically significant, invalidating the conclusions of the paper. Dr. Smith retracts this paper.

This is a retraction due to an honest error and represents a correction of the scientific literature. Dr. Smith may be sloppy, but he is not guilty scientific misconduct.

- f. A graduate student in the Brown lab is studying the impact of prairie dog colonization on ground water. The student uses remote sensors to measure ground water pH in several prairie locations daily, over a three month period. An analysis of these data reveals that two measurements are more than eight standard deviations from the mean. The student reanalyzes the data without these two data points and writes up the results. In the methods section, she states that two data points were dropped from the analysis and gives the statistical rationale for this decision.

Dropping these data points was justified because they represented a small subset of the total data, and because there was a statistical basis for identifying the dropped data as outliers. Further, in the manuscript, the author mentioned the dropped data points and her rationale for eliminating them from the analysis.

- g. Dr. Jones, of the College of St. Radegund, Cambridge University, is a very famous scholar in the area of rodent systematics. Dr. Smith sends Dr. Jones a copy of an unpublished manuscript on prairie dogs. Dr. Jones responds with positive comments and suggests a few minor changes in wording. Dr. Smith makes the suggested changes and submits the manuscript to a journal, after adding Dr. Jones to the author list.

This is inappropriate authorship. Dr. Jones has not contributed enough to the manuscript to merit authorship. Perhaps Dr. Smith thinks that the article will get more attention with Jones listed as an author.

- h. In the prairie dog paper, Dr. Smith introduced a novel phylogenetic method developed in her laboratory. Later, Dr. Jones uses this new method in an analysis of the evolutionary origins of chipmunks and writes up the results for publication. In Jones' manuscript, Dr. Smith's work on prairie dogs paper is cited in a section reviewing prior work on rodent evolution. Jones' methods section points out that the method used is novel, but does not state that it was first proposed by Dr. Smith's group.

Plagiarism. By leaving the reader with the impression that Jones invented the method, Jones is stealing ideas.

7. There are many contexts in which it is important to write a review or summary of prior research: in the background section of a grant proposal, in the introduction to a publication reporting the results of original research, or in a literature review article. When writing a review of this type, the author must take care to paraphrase appropriately; that is, to give a summary that *correctly* describes the original work and is written in the words of the reviewing author, without "borrowing" from the original publication.

In the following questions, you are asked to compare an excerpt from a publication with a paragraph paraphrasing the original text. Did the author of the second text summarize the original text correctly and appropriately? Did the author of the second text plagiarize the original text? Did the author of the second text paraphrase the original text incorrectly; that is, did she use her own words, but change the meaning of the original text? Explain your reasoning.

Original text: "Concluding remarks and perspectives. In conclusion, we present evidence that two independent engineered strains (ASI and COS) expressing high levels of KinA induce sporulation efficiently irrespective of nutrient availability. With ASI, we show that a small increase in the level of the KinA protein beyond a certain level triggers sporulation. Furthermore, we demonstrate that the COS strain bypasses the requirement of nutrient starvation conditions to trigger sporulation. In the COS strain, grown under conditions of nutrient excess, several key sporulation genes show temporal and spatial expression patterns, which are normally observed for the wild-type strain only under starvation conditions. These results, obtained from both ASI and COS cells, suggest that the threshold level, but not the signal-dependent activation, of the major sporulation kinase KinA acts as a molecular switch in the initial decision of whether to commit to sporulation or not, although the mechanisms that increase the level of KinA remain elusive."¹

New Text: Easwaramoorthy and colleagues (Easwaramoorthy P. et al. J. Bact. 192, 3870–82, 2010) present evidence that two independently engineered strains that express high levels of the major sporulation Kinase, KinA, are capable of efficient sporulation, irrespective of nutrient availability. In one strain, they showed that when KinA expression exceeds a threshold level, due to a small increase in abundance, KinA acts as a molecular switch in the initial decision of whether to commit to sporulation. In the other strain, they observed several temporal and spatial expression of key sporulation genes when the strain was grown under conditions of nutrient excess. In wild-type, these expression patterns are normally observed only under starvation conditions. Taken together, these results suggest that signal-dependent activation is not the deciding factor. The mechanisms that increase the level of KinA remain unknown.

This is verbatim plagiarism. Entire phrases are reproduced verbatim (see underlining above). This is unacceptable, even if the original work is cited. Although the writer mentions Easwaramoorthy et al. and does not attempt to present this work as his/her own results, this still constitutes plagiarism. The writer is presenting the words of Easwaramoorthy et al. as though they were his/her own.

¹ Eswaramoorthy P. et al. (2010) The Threshold Level of the Sensor Histidine Kinase KinA Governs Entry into Sporulation in *Bacillus subtilis*. *Journal of Bacteriology* 192, 3870–3882

8. Compare the following excerpt from a publication with the paragraph paraphrasing the original text. Did the author of the second text summarize the original text correctly and appropriately? Did the author of the second text plagiarize the original text? Did the author of the second text paraphrase the original text incorrectly; that is, did she use her own words, but change the meaning of the original text? Explain your reasoning.

Original text: "Concluding remarks and perspectives. In conclusion, we present evidence that two independent engineered strains (ASI and COS) expressing high levels of KinA induce sporulation efficiently irrespective of nutrient availability. With ASI, we show that a small increase in the level of the KinA protein beyond a certain level triggers sporulation. Furthermore, we demonstrate that the COS strain bypasses the requirement of nutrient starvation conditions to trigger sporulation. In the COS strain, grown under conditions of nutrient excess, several key sporulation genes show temporal and spatial expression patterns, which are normally observed for the wild-type strain only under starvation conditions. These results, obtained from both ASI and COS cells, suggest that the threshold level, but not the signal-dependent activation, of the major sporulation kinase KinA acts as a molecular switch in the initial decision of whether to commit to sporulation or not, although the mechanisms that increase the level of KinA remain elusive."²

New text: Easwaramoorthy and colleagues (Easwaramoorthy P. et al. J. Bact. 192, 3870–82, 2010) used two artificially constructed strains to investigate the conditions that govern the switch-like behavior of KinA, the signaling kinase that controls sporulation initiation. In one of these strains, they observed that a small increase in expression is enough to initiate sporulation. In the other strain, they observed sporulation initiation in response to environmental conditions very different from those that trigger sporulation in wild-type cells. Instead of responding to nutrient starvation, the mutant strain expresses key sporulation genes in response to nutrient excess! By decoupling KinA expression and environmental signals, Easwaramoorthy *et al.* were able to determine that KinA expression, and not environmental changes, are responsible for the decision to sporulate. The processes whereby KinA expression changes in response to environmental cues are still unknown.

The new text appropriately paraphrases the original text. Note that some two word phrases, such as "nutrient starvation" have taken on specific technical meanings in this field. These appear in both texts because these two-word terms have become part of the standard vocabulary used to describe this type of experiment.

² Easwaramoorthy P. *et al.* (2010) The Threshold Level of the Sensor Histidine Kinase KinA Governs Entry into Sporulation in *Bacillus subtilis*. *Journal of Bacteriology* 192, 3870–3882

9. Compare the following excerpt from a publication with a paragraph paraphrasing the original text. Did the author of the second text summarize the original text correctly and appropriately? Did the author of the second text plagiarize the original text? Did the author of the second text paraphrase the original text incorrectly; that is, did she use her own words, but change the meaning of the original text? Explain your reasoning.

Original text: "One set of circumstances under which some adaptive mutations can be prohibited is in the case of multifunctional genes. The performance of multiple functions by a single gene or 'gene sharing' is prevalent in nature but presents the possibility that mutations that optimize one function may compromise the other. Such an 'adaptive conflict' may be resolved by gene duplication followed by subfunctionalization and positive selection, thereby creating two optimized specialist genes in a manner forbidden during gene sharing."³

New text: In their 2007 Nature paper, Hittinger and Carroll (Nature 449, 677-682, 2007) discuss the fate of a multifunctional gene after gene duplication. There are many examples in nature of a single gene that performs two or more functions. The possibilities for evolutionary change in this case are limited, because a mutation that improves the efficiency of one of the functions may decrease the efficiency of the other function. Gene duplication may free such a gene from this "adaptive conflict." If each of the two gene copies only performs one of the two ancestral functions, then both copies can mutate to improve the efficiency of their respective functions, without compromising the function performed by the other copy.

The new text appropriately paraphrases the original text. Here the writer is paraphrasing for a more general audience, expressing the same idea, but avoiding technical terms. The structure of both texts is similar, but the new text expresses the ideas using phrases written in standard English for the non-specialist. For example, "each of the two gene copies only performs one of the two ancestral functions" is used instead of the term "subfunctionalization."

³ Hittinger C.T. and Carroll, S.B. (2007) Gene duplication and the adaptive evolution of a classic genetic switch. Nature 449, 677-682

10. Compare the following excerpt from a publication with the paragraph paraphrasing the original text. Did the author of the second text summarize the original text correctly and appropriately? Did the author of the second text plagiarize the original text? Did the author of the second text paraphrase the original text incorrectly; that is, did she use her own words, but change the meaning of the original text? Explain your reasoning.

Original text: "In this work, we have tracked the evolution of combinatorial gene regulation by the highly conserved transcriptional regulator Mcm1 and each of its known cofactors across the ascomycete fungal lineage. Our analysis shows that the genes regulated by Mcm1 have changed considerably over the evolutionary time scales represented by this lineage; our results reveal many more differences than similarities in the Mcm1 circuitry. Regulation by Mcm1 is more pervasive in *K. lactis* and *C. albicans*, where 12% of all genes are bound, than in *S. cerevisiae*, where 4% of genes are bound. The fraction of genes shared as targets between all three species is very low (13%–18%), and we have demonstrated that this is due to both substantial gain and loss of Mcm1 binding sites along each branch of this phylogeny ..." ⁴

New text: Tuch and his colleagues (Tuch et al, PLoS Biol., 6, 2008, p. e38) showed that genes regulated by the conserved transcription factor Mcm1 have changed considerably in the course of ascomycete fungal evolution. In fact, there are more differences than similarities in the MCM1 circuitry of the species studied. MCM1 regulates more genes in *K. lactis* and *C. albicans* than it does in *S. cerevisiae*. Moreover, there is little overlap between the targets of MCM1 in the three species. This is because on each branch of the fungal tree, a substantial number of genes have lost MCM1 binding sites, while a substantial number of other genes have gained MCM1 sites.

This is mosaic plagiarism. Although the phrases that are reproduced verbatim (see underlining above) are shorter and fewer in number than in problem 6, this level of reproduction of the original text is unacceptable.

⁴ B.B. Tuch, D.J. Galgoczy, A.D. Hernday, H. Li, A.D. Johnson (2008). The evolution of combinatorial gene regulation in fungi. PLoS Biol., 6 (2008), p. e38.

11. Compare the following excerpt from a publication with the paragraph paraphrasing the original text. Did the author of the second text summarize the original text correctly and appropriately? Did the author of the second text plagiarize the original text? Did the author of the second text paraphrase the original text incorrectly; that is, did she use her own words, but change the meaning of the original text? Explain your reasoning.

Original text: "In this work, we have tracked the evolution of combinatorial gene regulation by the highly conserved transcriptional regulator Mcm1 and each of its known cofactors across the ascomycete fungal lineage. Our analysis shows that the genes regulated by Mcm1 have changed considerably over the evolutionary time scales represented by this lineage; our results reveal many more differences than similarities in the Mcm1 circuitry. Regulation by Mcm1 is more pervasive in *K. lactis* and *C. albicans*, where 12% of all genes are bound, than in *S. cerevisiae*, where 4% of genes are bound. The fraction of genes shared as targets between all three species is very low (13%–18%), and we have demonstrated that this is due to both substantial gain and loss of Mcm1 binding sites along each branch of this phylogeny ..."⁵

New text: Tuch and his colleagues (Tuch et al, PLoS Biol., 6, 2008, p. e38) studied evolutionary changes in regulation by the yeast transcription factor MCM1 and its co-factors. Their analysis shows significant change in MCM1 regulation; in fact, expression patterns of MCM1-regulated genes are quite different in the species studied. However, the results could be compounded by the substantial gain and loss of genes in those species.

There is no plagiarism here, however the paraphrasing is incorrect:

- ***The original text states that the genes that are regulated by MCM1 have changed. The new text states that the expression patterns of genes that are regulated by MCM1 have changed.***
- ***The original text states that MCM1 binding sites have been gained and lost. The new text states that genes that have MCM1 binding sites have been gained and lost.***

The meaning in the new text is very different from the meaning in the original text.

⁵ B.B. Tuch, D.J. Galgoczy, A.D. Hernday, H. Li, A.D. Johnson (2008). The evolution of combinatorial gene regulation in fungi. PLoS Biol., 6 (2008), p. e38.