Measuring Scholarly Discourse Change with Respect to Citations – A Nobel Prize Case Study

Jessica Cox, Curt Kohler, Anthony Scerri, Corey Harper, Paul Groth, Ron Daniel, Jr.
{j.cox; c.kohler; a.scerri; c.harper; p.groth; r.daniel}@elsevier.com

Elsevier Labs, 1600 John F. Kennedy Boulevard, Suite 1800, Philadelphia, PA 19103-2822

Hypothesis

Citation behavior and terminology used will differ before and after the Nobel prize is awarded. We expect to see differences in terminology around the citation, differences in sections where the citation occurs and diversity in subject area to change after the prize is awarded.

Method

1. Identify all Nobel Prize awards from 1995-2017 in economics, medicine, chemistry and physics.
2. Select a paper(s) that encapsulates the discovery for which the prize was awarded, using nobelprize.org as a guide.
3. Compile the papers and their respective identifiers (authors, ScopusIDs, year of publication, etc.)
4. Search our XML catalog for full-text articles that cited this list of source papers.
5. Capture these citing articles, and their accompanying features, such as year of publication and subject area of the citing work.
6. Use Annotation Query to capture the sentence that contains a citation of one of the source papers.
7. Exploratory analysis of data:
   1. Group sentences based on their domain
   2. Group sentences relative to time of prize award, before or after
   3. Group sentences relative to placement within the citing document
8. Construct vocabulary of "seminal" terminology and asses their presence in various cross-sections, as indicator of change in discourse.

Data Selection

- **160** • Nobel Prize proxy papers (NPP)
- **335,187** • Papers indexed in Scopus that cited NPPs
- **39,096** • Papers as full-text XML in ScienceDirect that cited NPPs
- **75,432** • Sentences extracted from full-text XML that contain direct citation to NPP
- **16,858** • Sentences that contain a direct citation to NPP and contain a "seminal word"

References

Bjork, S. & Offer, A. & Söderberg, G. (2014). Time series citation data: The Nobel Prize in economics. Scientometrics 98(1). 10.1007/s11192-013-0989-5.

Chen, C., & Lobo, N. (2006). Analyzing and Visualizing the Dynamics of Scientific Frontiers and Knowledge Diffusion. *Encyclopedia of Human-Computer Interaction*, 24-30.

Cox, J. & Kohler, C & Groth, P (2018). Citation Contexts for Nobel Prize Wining Papers. Mendeley Data, v1. http://dx.doi.org/10.17632/g75gcpp49k.1

Jha, R. & Jbara, A. & Qazvinian, V., & Radev, D. (2017). NLP-driven citation analysis for scientometrics. *Natural Language Engineering*, 23(1), 93-130. 10.1017/S135132
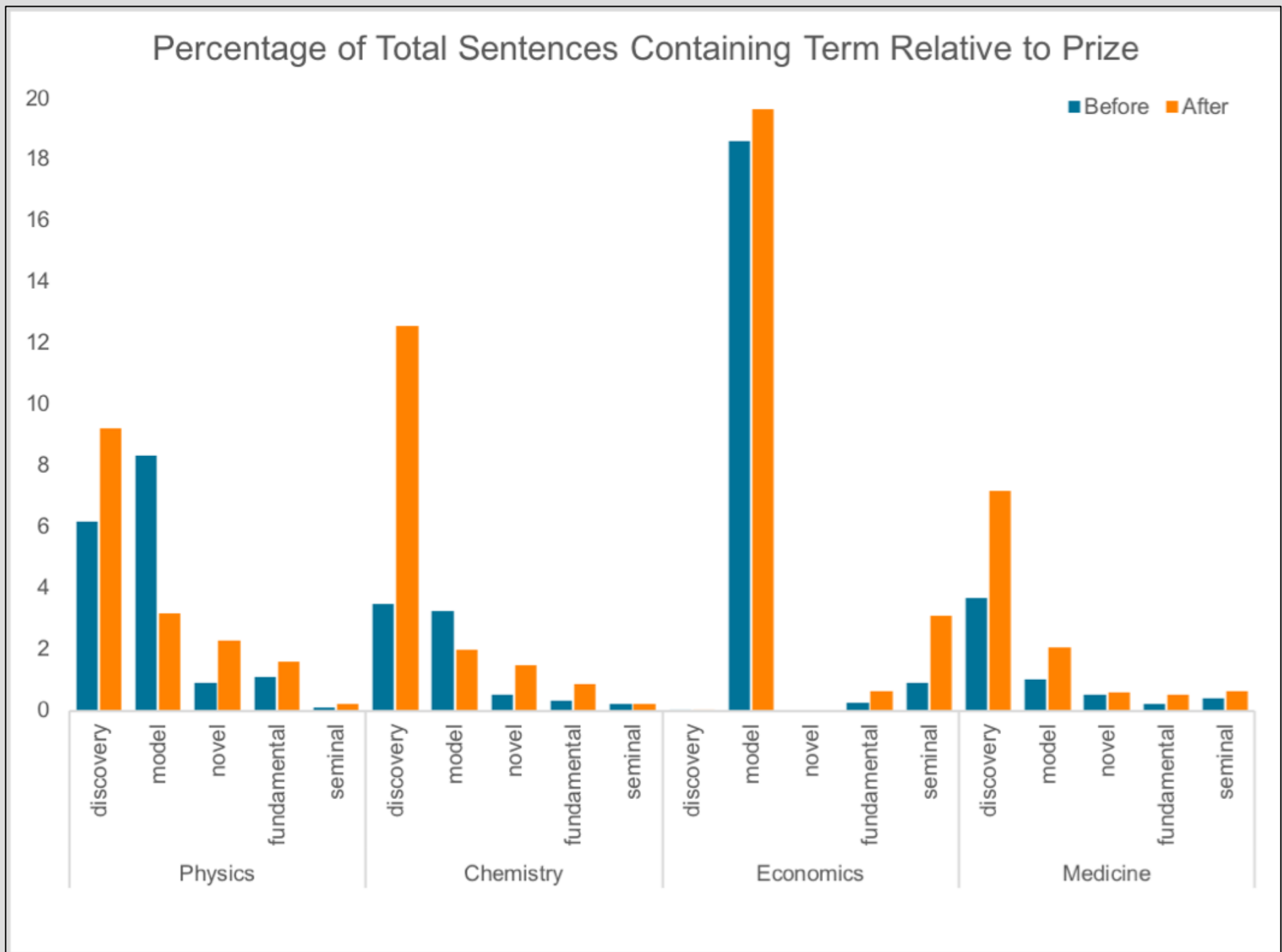
Conclusions

There were modest increases in specific "seminal words" used in describing a NPP after the prize was awarded. Additionally, we saw a shift in description of NPPs to occur primarily in the introduction, which had the highest occurrence of "seminal words". We also observed that the first time one of these words was used in a citation was consistently before the prize was awarded.

Results

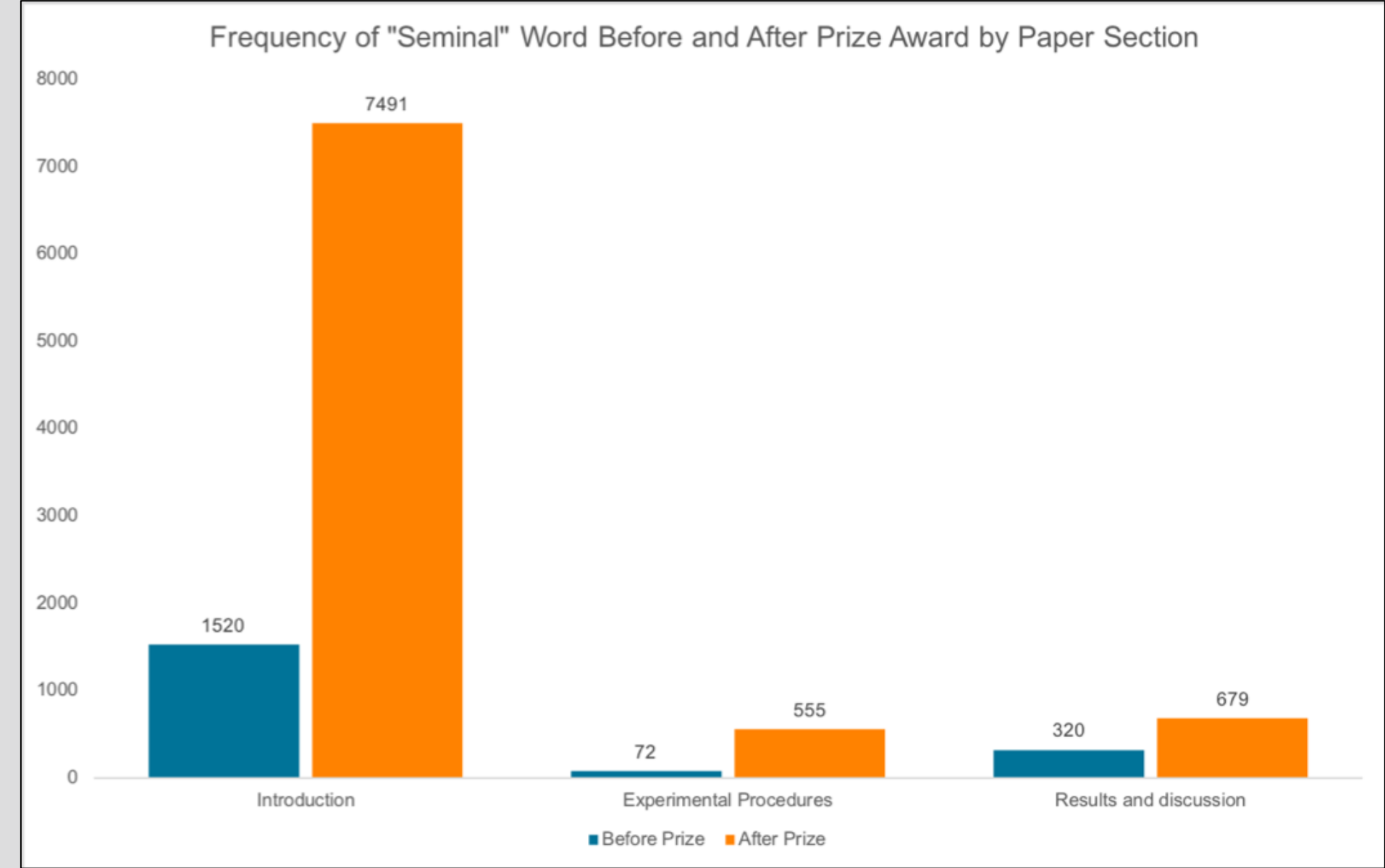Sentence Count Relative to Prize by Domain
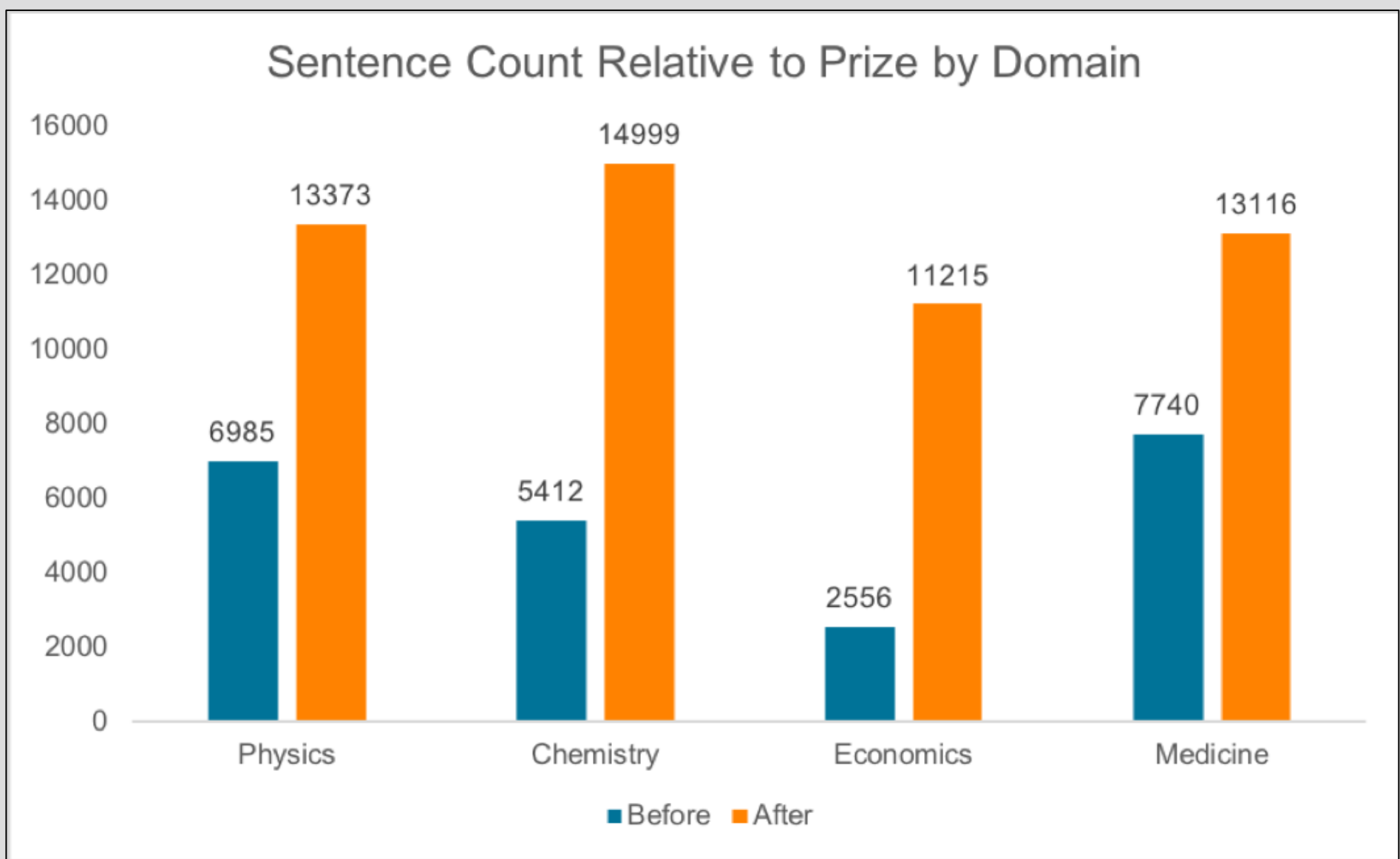
70% of the sentences occurred after the prize was awarded for its respective cited NPP.

22% of the sentences contained a "seminal word". For each NPP, the median time to first occurrence of a one of these words in their citing sentence was 4 years before the prize was awarded.

Percentage of Total Sentences Containing Term Relative to Prize

Specific terms were more prominent in some domains, such as model in economics. The word "discovery" increased after the prize for three of the domains. Overall, modest changes in word frequencies were observed.

Frequency of "Seminal" Word Before and After Prize Award by Paper Section

The use of "seminal" terminology was five times greater in the Introduction section after the prize was awarded to the cited NPP.