A Compromised Knowledge Base: Academic Publishing in Crisis

Hybrid Open Access: Monetizing Freedom of Knowledge

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Systems Engineering

By
Anna Fisher

October 7, 2024

Technical Team Members:
Ford Downer, Sean Ferguson, Riley Tomek

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Gerard Fitzgerald, Department of Engineering and Society

William Scherer, Department of Systems Engineering

Matthew Bolton, Department of Systems Engineering

Introduction

Academic research is the backbone of our world knowledge base as well as the driving force behind the massive academic publishing industry. Research journals have become the primary method of scholarly communication, with the University of Virginia's library spending 83% of its budget on journals (B. Butler, 2017). However, the rise of journals has not been without controversy. This research will focus on two issues the research community faces: questionable research practices through the technical project and accessibility through sociotechnical research. With rising pressure on researchers to publish in an academic world dominated by journals, some have resorted to undesirable methods as seen in a study in 2024 which estimates that one in every seven papers contain "errors consistent with faking" in at least one area (Heathers, 2024). These untrustworthy results infiltrate the academic knowledge base and compromise the whole system, putting scientific progress in jeopardy. In collaboration with RAND Corporation, the technical project aims to design a taxonomy detailing the types of academic fraud present in the publishing world. This will be used to develop a graph modeling tool to cluster papers, thus identifying key factors that flag a paper as a potential fraud.

The sociotechnical research will explore the development of hybrid open access through the Social Construction of Technology (SCOT) framework. Open access, a movement seeking to make literature "online" and "free of charge" without restrictions on "copyright and licensing", was formed in opposition to the exclusivity of commercial academic publishing (Suber, 2012, p. 4). The academic publishing industry has an oligopolistic nature with five companies controlling 61% of the market: Elsevier, Wiley, Taylor & Francis, Springer, and Sage (Crotty, 2023). This market share enables publishers to raise prices of journal articles, leaving libraries to shoulder the financial burden. Publishers charge hefty fees for subscription services, causing the UVA

library to spend 43% of their budget on deals with four publishers alone (B. Butler, 2017). With open access threatening their industry, publishers developed a type of publishing labeled Hybrid Open Access (Hybrid OA). Hybrid OA journals publish an article as open access if the author pays a fee; otherwise, the journal remains subscription-based (Suber, 2012, p. 140). This system, rooted in the exclusivity of knowledge, leads to large profits for publishers at the expense of institutions and researchers themselves. The SCOT framework will analyze the creation of hybrid open access through the lens of the affected social groups: researchers and authors, the top commercial publishers, libraries, and the United States government. Investigating questionable research practices and the accessibility of the publishing system will offer insight into the future of the world's knowledge base.

A Compromised Knowledge Base: Academic Publishing in Crisis

Corruption seeps into the academic knowledge base through many different avenues, some harder to detect than others. In one high profile example, researcher Scott Reuben faked at least 21 articles surrounding anesthesiology, many of which informed standard medical practice among doctors. Reuben pioneered methods in anesthesiology only for an internal investigation to provide evidence of falsified data and forgered coauthors (“Fraud Case Rocks Anesthesiology Community,” 2010). Reuben’s actions brewed mistrust towards anesthesiology practices while also directly affecting the civilian public. Unfortunately, it often takes years to detect research fraud, and even when detected, many cases of fraud remain hidden from the public. The blog Retraction Watch, dedicated to documenting retractions in academic literature, argues that retractions are not often “well-publicized” and journals need to be held accountable for the ways

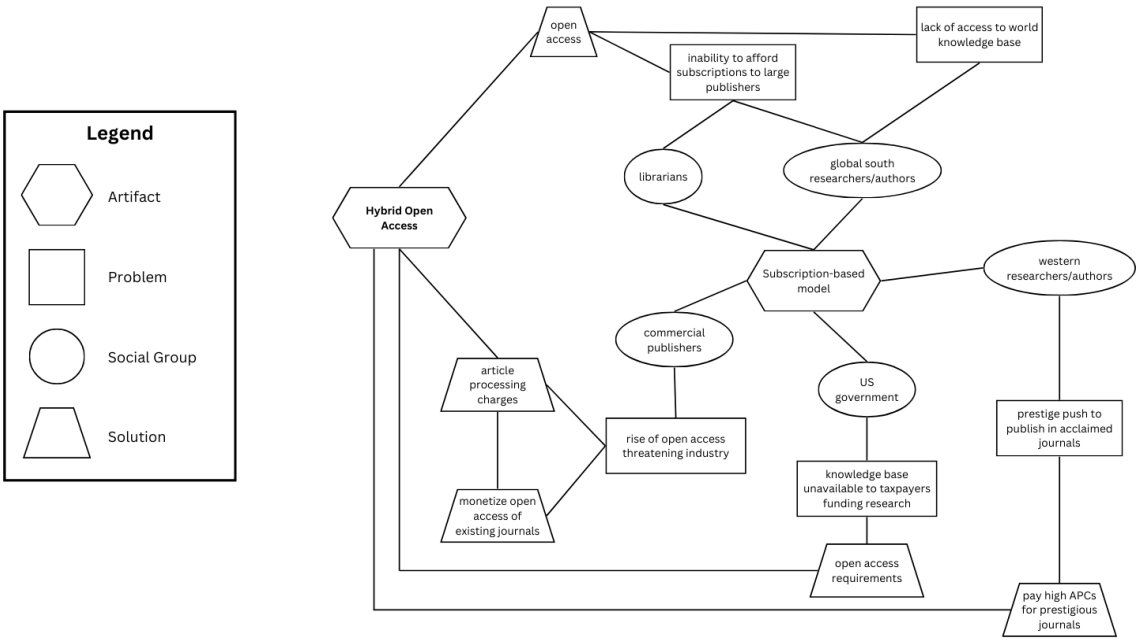
in which they handle retractions (Oransky & Marcus, 2010). Additionally, fraudulent behavior extends into multiple aspects of research. One researcher benefitted from over three thousand faked citations hidden in article metadata (Besançon et al., 2024). In a world where citation count can mean the difference between tenure and unemployment, false citations hurt honest researchers and elevate fraudsters to unearned stature.

While individual researchers are responsible for fraudulent actions, the underlying structure of the publishing system creates incentives for researchers to engage in questionable behavior. A scientist's career depends on being published, resulting in intense pressure to publish meaningful results at all costs (Buranyi, 2017). The "publish or perish" mindset drives researchers to seek citations and publications, putting the search for knowledge second to the need for prestige (Crous, 2019). Though the technical project focuses on the output of fraud into the academic community, the forces driving the demand for fraudulent research will also be examined. First, the research aims to deliver a full taxonomy of questionable research practices in academic publishing. Through a literature review and conversations with experts in the library science field, types of fraud will be classified in the form of a hierarchical taxonomy. This taxonomy will illustrate dependencies and quantify impact of the practice on the academic community. The taxonomy will also depict the financial incentives existing within academia to offer a full picture of the system enabling fraud. In the second phase of research, an artificial intelligence model will define dimensions of fraud, identifying key indicators of questionable practices within research papers. This tool will serve as a flag for fraudulent behavior, capturing a specific part of the fraud taxonomy. The goal of the technical project is to centralize information on questionable practices in research, explore incentives fueling this behavior, and develop a tool to detect fraud.

Hybrid Open Access: Monetizing Freedom of Knowledge

The sociotechnical aspect of this research will explore how researchers, librarians, the commercial publishers of Elsevier and Springer, and the US government impacted the evolution of the hybrid open access journal model using the SCOT framework. This framework explores the development of technology through the lens of affected social groups, as seen in Figure 1.

Figure 1: SCOT Framework



Hybrid open access refers to a type of publication that allows individual researchers to pay additional fees to publish their articles to the general public while the remainder of the journal remains behind a paywall. With journal articles becoming the main medium for sharing knowledge, controlling access to scientific publications is like controlling science itself (Buranyi, 2017). Thus, understanding the mechanisms for journal distribution has significant implications for the future of science. As large commercial publishers continue to monetize research funded

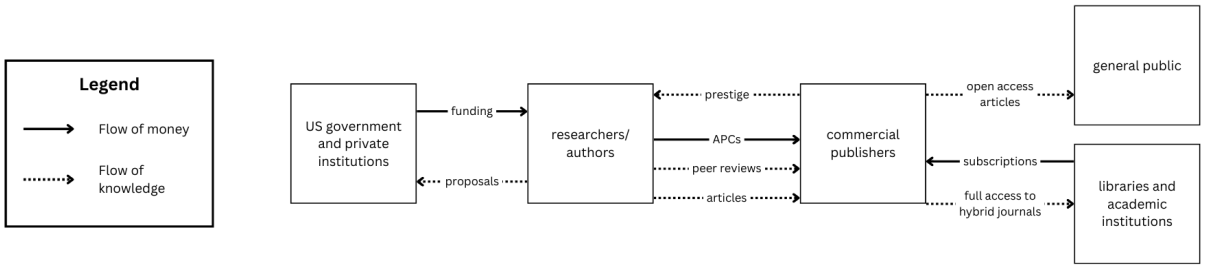
by the government with minimal financial gain to authors, the distribution of science impacts progress worldwide.

The commercial publishing industry originated with the subscription-based model: publishers rely on submissions from separately-funded academics, sourcing voluntary peer reviews before charging libraries (and their researchers) millions of dollars in subscription fees for journal access (Buranyi, 2017). Before the internet age, publishers incurred large costs in procuring peer reviews and distributing hard copies of journal issues, justifying high subscription prices. However, with electronic copies of journals easing both the peer review and distribution costs, publishers enjoyed unprecedented profit margins while journal subscriptions remained high, often at the expense of the same academics providing peer reviews and research free of charge. For example, in 2023, Elsevier's parent company reported an astronomical profit margin of 33.1% compared to the 12-15% profit margins collected by successful non-academic magazines (RELX 2023 Results, 2024; Buranyi, 2017). These profit margins are made possible by limiting access to knowledge, creating artificial scarcity through paywalls. Though a subscription to one journal may be a trivial cost, paying for the large volume of work published each year becomes "insurmountable", creating an "access gap" (Suber, 2012, p. 4). Publishing companies have historically profited off of the exclusivity of research and volunteered labor of the academic community through the subscription-based model, resulting in barriers to academic knowledge.

In response to the growing issue of paywalls surrounding knowledge, the open access movement was born. Originally created to oppose barriers to academic knowledge built on the free labor of researchers, open access releases journals from paywalls in exchange for a fee called an Article Processing Charge (APC) paid by the author (Suber, 2012, p. 9). With librarians

desperately searching for a solution to colossal subscription fees and authors seeking freedom of knowledge, open access seemed poised to render the traditional commercial publishing industry obsolete. Instead, publishers introduced the hybrid open access journal format, cementing their market share and successfully monetizing open access (Shu & Larivière, 2024). Hybrid open access generates substantial revenue by collecting fees from both authors and libraries, shown in Figure 2. Journals that are not fully open access, still operating under the subscription-based

Figure 2: Hybrid Open Access Flowchart



model, offer authors the option of publishing their articles open access at the expense of an APC. While the article still appears under the name of a subscription-based journal, the article itself will not be behind a paywall. Though based on the original open access model, publishers charge libraries additional subscription fees to access the remainder of the journal, receiving compensation from both authors and academic institutions. The charges commercial publishers require in their hybrid journals are no trivial amount. From 2015 to 2018, the five major publishers garnered $448.3 million in APCs from hybrid journals (L. Butler et al., 2023). Often, researchers must rely on institutions or grant funding to accommodate these high fees; at worst, researchers must pay out-of-pocket. With institutions' use of prestigious publications to judge promotions and tenure, authors find themselves choosing to either pay high APCs for prestigious hybrid journals or risk their career. Publishers use APCs to position themselves as key financial

beneficiaries in open access, a system formed with the goal of removing financial barriers from research.

The US government recently required all federally funded research to be published open access through the 2022 Nelson mandate in an attempt to make knowledge accessible to the taxpayers responsible for funding (Winter, 2022). While the open access movement swells in importance, this research investigates the development of the hybrid open access model to inform open access policy for the future. By conducting a literature review and utilizing experts in library science, this analysis will illustrate how social groups of librarians, commercial publishers, the US government, and researchers influenced the creation of the hybrid open access journal. This project will utilize the publishers Elsevier and Springer Nature as case studies due to their extensive profits from hybrid open access journals (L. Butler et al., 2023). Using the SCOT framework, this research will investigate the underlying incentives of academia causing librarians and researchers to participate in a financially exploitative publishing model. Thus, this project will attempt to analyze the social dynamics that produced the hybrid open access system.

Conclusion

The capitalist nature of the publishing industry paired with intense pressure on academics creates many incentives for fraud. This high influx of unreliable information contaminating the world knowledge base has implications for technological progress today and into the future. Untrustworthy research produces an unstable knowledge base where nothing can be trusted, and thus nothing can serve as a basis for progress. The technical project with RAND Corporation will deliver a taxonomy of questionable research practices occurring within the academic community

and an artificial intelligence tool to identify fraud in a determined area, attempting to answer how and why fraudulent research prevails in the academic knowledge base. By understanding the scope of fraud and defining fraudulent practices, this research will serve as a foundation for further work on the integrity of academic literature.

The sociotechnical element of this research will deliver a SCOT analysis of the formation of the hybrid open access publishing model. Publishers continue to profit off of the work of authors and cultivate an environment of academic exclusivity through hybrid open access journals. These journals allow publishers to monetize a system built on the freedom of knowledge. This research aims to understand the incentives of librarians, authors, commercial publishers, and the United States government that drove the creation of the hybrid open access system. Exploring the development of hybrid open access through the lens of social groups will offer insight into the profit-driven publishing industry. Establishing a foundational understanding of questionable practices in research and investigating a profitable commercial open access model contributes to the movement towards a knowledge base free from misinformation and exploitative publishing.

References:

Besançon, L., Cabanac, G., Labbé, C., & Magazinov, A. (2024). "Sneaked references: Fabricated reference metadata distort citation counts." *Journal of the Association for Information Science and Technology*. pp. 1–12.

Buranyi, S. (2017, June 27). "Is the staggeringly profitable business of scientific publishing bad for science?" *The Guardian*. Accessed on October 6, 2024 from https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science

Butler, B. (2017, March 22). *Predatory publishing: From "fake academia" to real extortion in scholarly publishing* [Audio recording]. https://avalon.lib.virginia.edu/media_objects/f1881k94d

Butler, L., Matthias, L., Simard, M., Mongeon, P., & Haustein, S. (2023). "The oligopoly's shift to open access: How the big five academic publishers profit from article processing charges." *Quantitative Science Studies*, *4*(4), pp. 778–799.

Crotty, D. (2023, October 30). *The Scholarly Kitchen*. Accessed on October 6, 2024 from https://scholarlykitchen.sspnet.org/2023/10/30/quantifying-consolidation-in-the-scholarly-journals-market/

Crous, C. J. (2019). "The darker side of quantitative academic performance metrics." *South African Journal of Science*, *115*(7/8). https://doi.org/10.17159/sajs.2019/5785

"Fraud Case Rocks Anesthesiology Community." (2010, October 8). *Anesthesiology News*. Accessed on October 6, 2024 from

https://www.anesthesiologynews.com/Policy-Management/Article/03-09/Fraud-Case-Rocks-Anesthesiology-Community/12634?ses=ogst

Heathers, J. (2024). “How Much Science is Fake?” *Center for Open Science*. https://doi.org/10.17605/OSF.IO/5RF2M

Oransky, I., & Marcus, A. (2010, August 3). “Why write a blog about retractions?” *Retraction Watch*. Accessed on October 6, 2024 from https://retractionwatch.com/2010/08/03/why-write-a-blog-about-retractions/

RELX 2023 Results. (2024). [Press Release]. RELX Group. https://www.relx.com/media/press-releases/year-2024/relx-2023-results

Shu, F., & Larivière, V. (2024). “The oligopoly of open access publishing.” *Scientometrics*, *129*, pp. 519–536.

Suber, P. (2012). *Open Access*. Cambridge, Massachusetts: The MIT Press. pp. 1-27, 133-142. https://muse.jhu.edu/book/46989

Winter, C. (2022, December 9). “The Nelson Memo: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research in the US.” *Open Scholarship Policy Observation*. https://doi.org/10.25547/2K93-PD42