Leaks, Attribution, and Academic Research*

The Role of Precedence and Normalization in Publication Trends

Thomas Fox

February 13, 2024

Ethical and professional issues face the research community while considering the millions of classified documents released by WikiLeaks as potential research materials. This paper examines trends in top-rated international relations journals that cite or make reference to these materials in publications between 2010 and 2020. Findings indicate an increase in articles published by peer-reviewed journals which use these materials, as well as a lack of professional consistency while doing so. This suggests the need for clear professional consensus within the research community surrounding these ethically questionable materials.

Table of contents

*Code and data are available at: https://github.com/ThomasWilliamFox/leaks_attribution_research.git;Replication on Social Science Reproduction Platform can be found at: https://www.socialsciencereproduction.org/reproductions/302459e5-ae78-49cf-82d1-265bfe0ad5f1/index

1 Introduction

First Paragraph. (Why it matters)

This paper follows a reproduction of Darnton's 2022 paper, "The Provenance Problem: Research Methods and Ethics in the Age of WikiLeaks" by examining its replication data made available through the Harvard Dataverse (Darnton 2021). The analysis of the data was carried out with the programming language R (R Core Team 2023). Various packages were used in the processing, cleaning, and presentation of this information, including tidyverse (Wickham et al. 2019), knitr (Xie 2023), here (Müller 2020), dplyr (Wickham et al. 2023), and janitor (Firke 2023).

Third Paragraph. (What is found)

This remainder of this paper contains an outline of the data examined, an analysis of the data, and a discussion of the findings. Section 2 outlines the analysis data made available by Darnton (Darnton 2021). Section 2.1 examines the TRIP-20 journals and their ratings Maliniak, Peterson, and Tierney (2012). Section 2.2 outlines the data-set built by Darnton, consisting of information about articles that use classified materials (Darnton 2021). An exploration and analysis of the data is found in Section 3. The paper concludes with a discussion surrounding the paper's findings, including potential weaknesses and further research Section 4.

2 Data

2.1 TRIP Survey International Relations Journals (2011)

The 2011 Teaching, Research, and International Policy (TRIP) faculty survey was published in 2012 as part of the TRIP Project at The Institute for the Theory and Practice of International Relations at the College of William & Mary in Williamsburg, Virginia (Maliniak, Peterson, and Tierney 2012). The survey, collected responses from 3464 political science, international relations, and foreign policy professionals across 20 countries (Maliniak, Peterson, and Tierney 2012). The table on page 52 of the publication is a world-wide aggregate of responses on two questions surrounding which journals have the "greatest influence on the way IR scholars think about international relations" (Maliniak, Peterson, and Tierney 2012). The top 20 journals from the TRIP survey were used the basis for Darnton's analysis data set of articles apparently referencing classified documents (Darnton 2022). Table 1, is built using data found within Darnton's replication data set (Darnton 2021). The data consists of the names of the 20 journals, their short form abbreviations used by Darnton, as well as their ranking in the TRIP survey, with 1 being most influential and 20 being the least.

Table 1: Names, Shortform, and TRIP-20 ratings of the 20 journals

Journal Name	Short Form	TRIP-20 Rating
International Organization	IO	1
International Studies Quarterly	ISQ	2
International Security	IS	3
Foreign Affairs	FA	4
American Political Science Review	APSR	5
World Politics	WP	6
European Journal of International Relations	EJIR	7
Journal of Conflict Resolution	JCR	8
Foreign Policy	FP	9
Review of International Studies	RIS	10
Millennium: Journal of International Studies	MIL	11
American Journal of Political Science	AJPS	12
International Affairs	IA	13
Security Studies	SS	14
Review of International Political Economy	RIPE	15
Journal of Peace Research	JPR	16
International Studies Review	ISR	17
International Relations	IR	18
Comparative Politics	CP	19
Global Governance	GG	20

2.2 Articles Apparently Referencing Materials From WikiLeaks (2010 - 2020)

A data set containing information related to articles apparently referencing classified materials released by WikiLeaks was built by Darnton as part of their paper (Darnton 2022). The terms, "Wikileaks", "cable", and a combination of the two were used to search for articles published between 2010 and 2020 in TRIP rated journals. Darnton used data bases such as JSTOR, Project Muse, and ProQuest, as well as publisher platforms including Sage, Wiley, and Oxford University Press to locate the articles contained in the data set (Darnton 2022). Of the 565 articles produced through their search, 397 were false positives (mentioned WikiLeaks organization or physical cables). Darnton read and itemized each of the 168 remaining articles to build their analysis data set (Darnton 2022).

Table 2 is built using Darnton's replication data (Darnton 2021). The data set contains 168 entries, each corresponding to an article in a TRIP rated journal that apparently references classified documents released by WikiLeaks. The "code" variable refers to Darnton's coding system, with "2" representing articles that refer to classified material via secondary sources, and "3" representing articles that quote, cite, or paraphrase classified material from WikiLeaks

directly (Darnton 2022). The "Journal" and "Year" variables refer to the journal that published the article and the year in which it was published.

Various Boolean variables (coded as "y" for yes and "n" for no) are usedin Darnton's data set. The "Classified" variable indicates whether the article makes reference to the classified nature of the materials (Darnton 2022). "No Source" indicates that at least one referenced document is missing a citation. "Leak" indicates that the article mentions leaked nature of sources. "Uses Euphemism" refers to articles that use verbs other than leak such as "publish, or disseminate." The "To-From-Found" variable describes whether citations include an "author (can be"embassy"), recipient (even "State"), and repository". "Peer reviewed" indicates whether the journal is peer reviewed."Cable Discussed" indicates whether an article discusses leaked US cables and the "Cables Cited" variable communicates how many individual leaked cables are cited.

Table 2: Sample of Cleaned Academic Publication Data

Code	Journal	Year	Mentions 'Classified'	No Source	Mentions 'Leak'
3	APSR	2020	n	n	n
3	IA	2020	n	n	n
3	IA	2020	n	n	n
3	IS	2020	y	n	n
3	ISR	2020	n	n	n
3	SS	2020	y	n	y

Uses euphamism	To-From-Found	Peer Reviewed	Cable Discussed	Cable(s) Cited
n	y	y	y	0
n	n	y	y	4
n	n	y	n	0
n	n	y	y	8
n	n	y	y	4
n	n	y	y	11

Figure 1 displays all 168 articles published in TRIP rated journals from 2010-2020 that apparently cite or reference leaked documents from WikiLeaks either directly or through secondary sources. The years with the greatest number of articles published are 2011 with 24, 2013 with 21, and 2016 with 20.

3 Results

Figure 2 displays

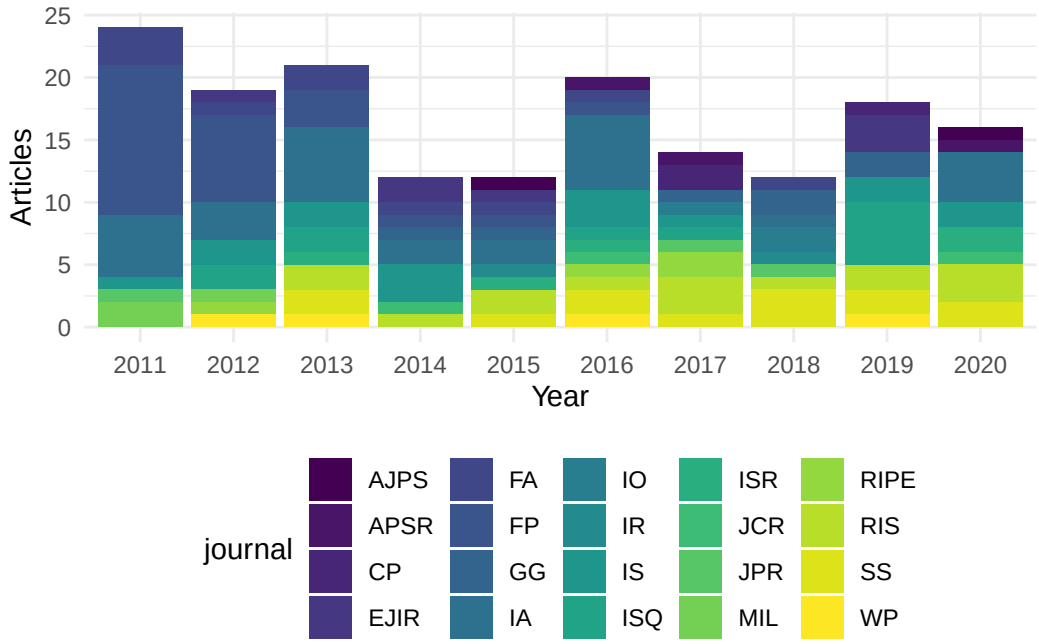

Figure 1: Articles in TRIP-20 rated journals apparently citing documents from WikiLeaks by year

Some of our data is about articles (**?@fig-fafpalarticles**).

Talk more about it.

And also planes. (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

4 Discussion

4.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

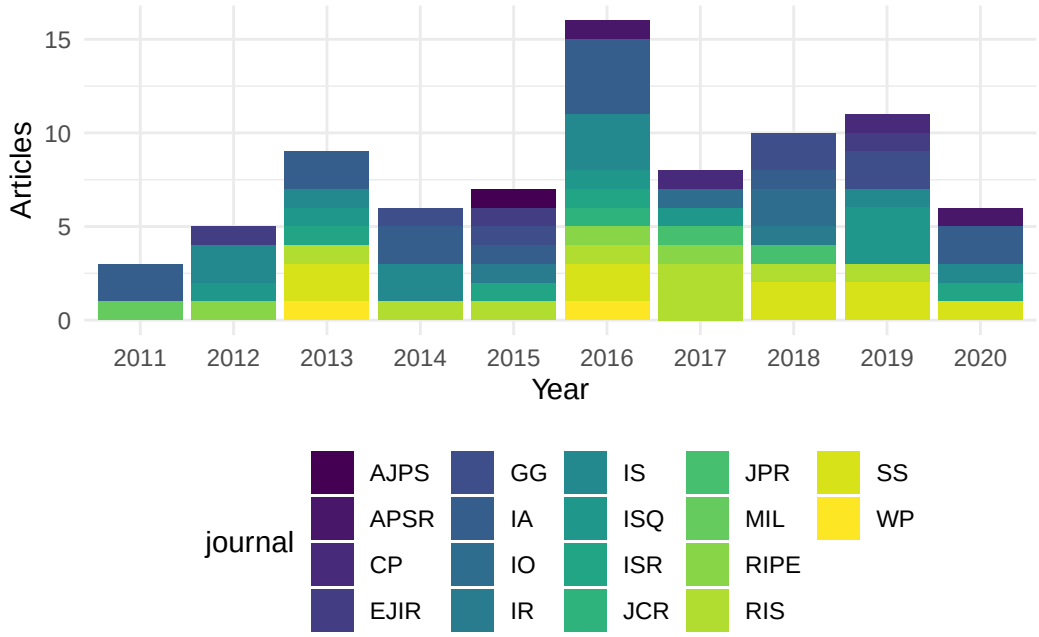

Figure 2: Articles in TRIP-20 rated journals apparently citing documents from WikiLeaks by year

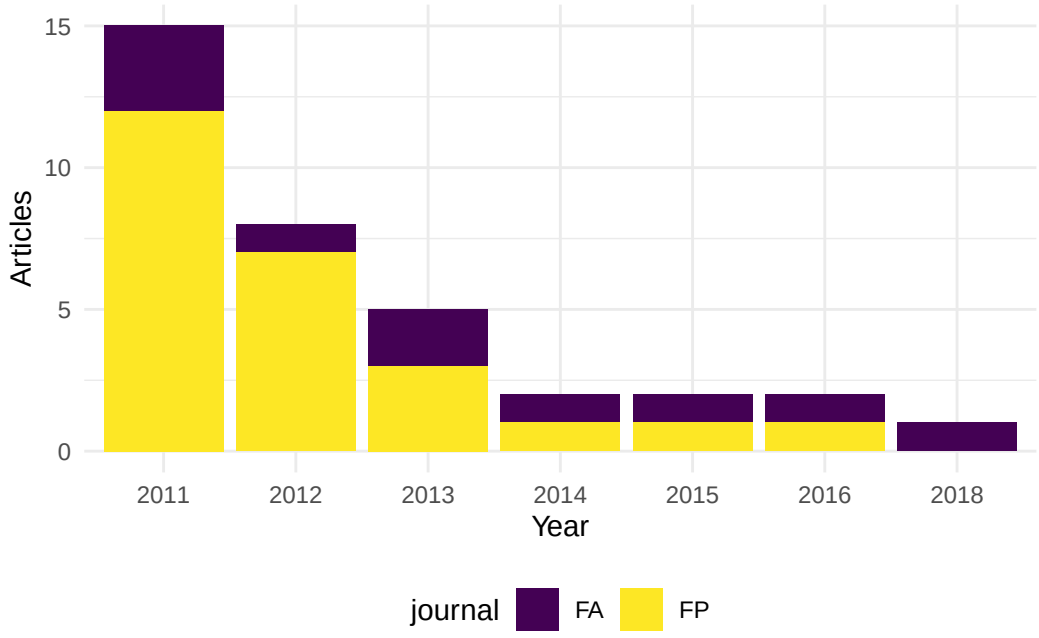

Figure 3: Articles in FA and FP apparently citing documents from WikiLeaks by year

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com.

Darnton, Christopher. 2021. *Replication Data for: The Provenance Problem: Research Methods and Ethics in the Age of WikiLeaks*. Harvard Dataverse. Dataset. https://doi.org/10.7910/DVN/SPLDTF.

———. 2022. "The Provenance Problem: Research Methods and Ethics in the Age of WikiLeaks." *American Political Science Review* 116 (3): 1110–25. https://doi.org/10.1017/S0003055421001374.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Maliniak, Daniel, Susan Peterson, and Michael Tierney. 2012. *TRIP Around the World: Teaching, Research, and Policy Views of International Relations Faculty in 20 Countries*. Williamsburg, VA: The College of William & Mary. https://www.wm.edu/ offices/itpir/_documents/trip/trip_around_the_world_2011.pdf.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.