

Spies, Cartels, Hackers, and Whistleblowers: Describing Those Who Do Not Want to Be Described

Brian Nussbaum, University at Albany

As has been noted (Von Maravic 2012), conducting research using traditional social science and public policy methods on organizations and individuals that actively hide their behavior is a challenge. Such organizations (intelligence services, mafias, cults, and criminal organizations) and individuals (hackers, assassins, fraudsters, and whistleblowers) do not lend themselves to traditional institutional and bureaucratic data creation and collection, engage in active attempts at deception and disinformation, and thus are hard to analyze using the traditional methods used to interpret such artifacts in aggregate.

This leads to the literature about such phenomena typically having many theoretical pieces and individual or small-N case studies, but limited comparative, large-N, or cross-national analysis. There have been some interesting advances and experiments in attempting to find new data - like court documents (Kim 2018) - or new methods - like social network analysis (Musotto 2022) - to address some challenges, but they remain incomplete, inchoate, or applicable only in cases where such data is publicly accessible.

The combination of efforts at obfuscation by the objects of study, which results in limited or misleading raw data, and the very low base rates of occurrence, means that – for at least the foreseeable future – description and case studies are likely to remain one of (if not the main) tools for studying such phenomena. This analysis intends to focus on why that, while imperfect, may actually be alright.

Many of the traditional methods applied to social science are based on 1) the idea of reliable data (or reasonable proxy data) and 2) the idea of generalization and cross case similarity; in the case of secretive subjects the data are likely to be limited and/or untrustworthy, and in the case of very low base rate activities (like whistleblowers or assassins) generalization may be less useful than many expect. Perhaps, “mere description” (Gerring 2012) is the right tool for many of these secretive and obscure phenomena.

Social Science vs. Intelligence Analysis: Generalizability and the Specificity of the Case

A social scientist wants one thing, and an intelligence analyst wants something else. A social scientist wants generalizable insights into social phenomena that are useful across many cases, and an intelligence analyst wants to leverage a mix of such generalizable insights and case-specific facts in order to offer the most reliable insights about an individual case. Not “What causes civil wars?” but “What is likely to push country X into civil war?” Both are useful, both are important, and to meaningfully understand any phenomena both approaches are required and can provide value. That said, there are phenomena for which the data and tools of each is better suited. Well-documented and transparent phenomena (GDP growth, legislative initiatives, level of economic development) are easier to study with the tools of social science

(and even this is - obviously - often very hard), than are those in which data is absent and untrustworthy and the actors attempt to deceive those examining them.

The “stigma of small N” looms large in the social science thinking around looking at hidden, covert, and illicit phenomena. But it is worth noting that when looking at hidden, covert and illicit phenomena, even small-N and single-N studies face huge challenges to be done effectively. When looking for generalizability across many cases, such stigma may make some sense, but in under-documented phenomena (or those with active mis/dis-information central to the data creation process), the disciplinarily driven push toward achieving large-N comparative studies may well lead to worse analytical outcomes.

Strategy 1) Process Tracing, Social Network Analysis (SNA), and Comparing Dynamics Across Related Phenomena

One approach that analysts could adopt would be to look at general dynamics across related, but different, phenomena. Using process tracing, social network analysis, and related case-level analyses - and then comparing them across both cases **AND** across related phenomena – is one potentially useful approach. This could mean, comparing social network data drawn from court case documentation across trafficking networks (drug, arms, human, and wildlife traffickers) to take the small-N (of “well documented” illicit networks) and aggregating up to slightly-larger-N studies of such phenomena.

Strategy 2) Tough Tradeoffs and the Narrow Utility of Generalizability

Another approach, and one I must admit being more partial to than many traditional social scientists, would be to accept that some phenomena are not well suited to large-N analysis, or at least are better suited to analyses driven by case studies or even theoretical approaches derived from more anecdotal data. The multi-method and mixed-method revolution has been valuable, and perhaps mostly because it was about the idea that the choice of methodological tools should be fit for purpose and fit the research question. The truth (sad truth, for some) is that some phenomena are not well suited to examination with the traditional tools of social science. The under-documented, undocumented, mis-documented, dis-documented, and otherwise “difficult access” problems are among this important set of problems.

Strategy 3) Multi-disciplinarity and Leveraging the Work of Non-Social Scientists

Finally, while in academia many incentives are set up to support research and publishing in individual disciplines, this strategy is exceptionally challenging with difficult access problems. It is not just academic disciplines that constrain the analysis of these problems, but also the fairly limited use of the work and methods of those studying such problems outside of academia. While academic research is used to drawing on the work of think tanks, more could be done to leverage the work of journalists, analysts at cybersecurity companies, and human rights activists who routinely document (albeit often anecdotally) the behavior of criminals, fraudsters, hackers, and spies.

Concluding Thoughts

Broad, standardized, and effective data collection, experimental design, hypothesis testing, and replication studies, are the *sine qua non* of social science, as it approaches traditional “hard” sciences. That said, the adoption of the expectation of those standards for phenomena that are fundamentally hard (or even... gasp... impossible) to effectively collect standardized and replicable data about is not just unreasonable, but also a sure way to guarantee unwarranted confidence and analytic missteps.

The call here would not be to stop trying for the goals of social science in these fields, pursuing the goals is worthwhile when it is reasonable, and the tools are fit for purpose. The call here is to not dismiss as “mere description” or “mere history” or “mere anecdote” the single and small-N studies that help to develop our baseline understanding of these very challenging phenomena.

Increasingly, the tools of “Open Source Intelligence” or OSINT are being harness to increase the documentation and data collection components of this problem. However, many of the most effective practitioners of this set of tools, have been organizations involved in non-social science and case specific analyses; investigative journalists (collectives like Bellingcat), intelligence agencies, and companies in the cybersecurity (CrowdStrike, Mandiant, FireEye, etc.) and private intelligence space (Janes, Strider, etc.). This is not because social scientists can’t make use of such information and data, but rather because it often lacks the structure and comparability of the data they seek. It’s certainly true that such data can inform data collection and data creation efforts, but it is often either 1) more useful at the case specific level or 2) sufficiently labor intensive to collect, that the most immediate and/or valuable analytic bang for buck comes instead from the smaller-N.

Gerring, J. (2012). Mere description. *British Journal of Political Science*, 42(4), 721-746.

Kim, A. C. (2018). Prosecuting Chinese Spies: An empirical analysis of the economic espionage act. *Cardozo L. Rev.*, 40, 749.

Musotto, R. (2022). *Understanding Mafia Networks*. Taylor & Francis.

von Maravic, P. (2012). Limits of knowing or the consequences of difficult-access problems for multi-method research and public policy. *Policy Sciences*, 45(2), 153-168.