MILLENNIUM
Journal of International Studies

Response

Millennium: Journal of
International Studies
2015, Vol. 43(3) 998–1002

DOI: 10.1177/0305829815581536
mil.sagepub.com
SAGE

Of Algorithms, Data and Ethics: A Response to Andrew Bennett[1]

Can E. Mutlu
Bilkent University, Turkey

Abstract

Developments in the field of Information and Communication Technologies (ICT) will have a significant impact on the way we study international relations. Opportunities related to data processing and automated reasoning that emerge through developments in complex algorithms will inevitably generate a debate on research methods in social sciences. Algorithms provide novel and innovative ways to sort and make sense of digital data. Applications of 'big data' and its potential uses in the social sciences remain understudied in IR. The field has not fully picked up on the potential uses of algorithmic processing for research. This article looks at the ethical questions that arise from the use of algorithmic data processing and automated reasoning. In particular, the article asks whether there should be any ethical limitations on the ways we collect data to be processed by algorithms.

Keywords

algorithms, data collection, ethics, privacy

Algorithms provide novel and innovative ways to sort and make sense of 'searchable and machine-readable' digital data. They are codes that instruct the machines to pursue a step-by-step set of operations to be performed automatically. The primary functions of algorithms are calculation, data processing and automated reasoning. Algorithms' significance to the social sciences in general, and to textual data processing in particular, is becoming especially clear in relation to the 'big data' revolution that seems to attract the interest of academics and practitioners alike. According to Andrew Bennett, the

1. Response to Andrew Bennett's keynote 'Found in Translation: Combining Discourse Analysis with Computer Assisted Content Analysis', 984–997.

Corresponding author:
Can E. Mutlu, Bilkent University Department of International Relations, Bilkent University Main Campus, A-253, Çankaya, Ankara, 06800, Turkey.
Email: canmutlu@bilkent.edu.tr

opportunities stemming from algorithmic methods for studying digital texts could bridge the divide between different epistemological traditions in International Relations (IR) scholarship. He might be right about that. The abilities stemming from algorithmic processing power are impressive and will undoubtedly have a profound effect on the research methods that we use in social sciences.

In this short response, I would like to focus on the failure to discuss the ethics of digital data collection in Bennett's article, in order to generate a productive debate on the uses of computer-assisted methods such as algorithmic data processing and automated reasoning. Bennett's emphasis on the transformative potential of algorithmic methods in relation to discourse analysis and computer-assisted content analysis omits a discussion of the ways in which it is ethical to collect large-scale data for algorithmic processing. As I argue in this intervention, any discussion of algorithmic data processing must include a discussion of data collection ethics. Given their processing power, researchers need large-scale, or 'big', data to harness the full potential of algorithms. Any attempt to collect mass-scale data without an ethical guideline will inevitably result in the breach of personal privacy and negatively affect the overall ethical integrity of the research process.

In his article 'Found in Translation: Combining Discourse Analysis with Computer Assisted Content Analysis', Bennett presents an argument that aims to bridge the gap between discourse analysis and computer-assisted content analysis by focusing on the impact of computer-assisted methods, including algorithmic processing. I believe that there is merit in discussing multi-method research and the uses of computer-assisted techniques such as algorithmic data processing and automated reasoning. However, as I argue in the following paragraphs, doing so without a clear discussion of the ethics of data collection presents the risk of infringement on the personal privacy of the unknowing human research subjects whose data is mined, used and analysed without their consent. With this intervention, I want to provoke a debate on the ethics of data collection and underline the increasing need for such a debate given the already ongoing 'algorithmic turn' in the social sciences in general, and IR in particular.

Bennett is right to suggest that developments in the field of Information and Communication Technologies (ICT) will have a significant impact on the way we study international relations. Algorithmic data processing and automated reasoning will and do allow scholars working on (inter)textual analysis to process a much greater number of texts, establishing more relationalities, across multiple media, and thus providing a more in-depth picture of the discursive terrain surrounding a particular issue. This does have the potential to bring different methods of textual analysis – discourse analysis and content analysis – closer to each other. The developments in the field of ICT in general, and algorithmic information theory in particular, allow us to collect/mine significant quantities of data. But more importantly, algorithms allow us to process data on a scale that was previously inconceivable. Given the automated reasoning capabilities of algorithmic processing which allow us to automatically sort through and establish relationalities within the extremely large pools of data, 'this algorithmic turn' represents a major shift in our capabilities as researchers. These developments will inevitably and undoubtedly change the reliability and accuracy of forecasting and modelling practices that are common in social sciences, as we will move away from 'sampling' and into monitoring and measuring almost real-time 'big data' to model and forecast in different and, perhaps,

more accurate ways through algorithmic processing. This analytical potential of algorithmic processing that makes it so attractive to policymakers and practitioners is also what makes it so attractive for social scientists.

On this point, Bennett's focus on the uses of algorithms and other digital methods is interesting and worth taking seriously. Bennett is right about the potential of algorithms and other machine-learning protocols for social science research for the reasons I have discussed above. Applications of 'big data' and its potential uses in the social sciences remain understudied in IR. The field has not fully picked up on the potential uses of algorithmic processing for research. There are, however, already existing research agendas in sociology in general, and surveillance studies and critical security studies in particular, that have been focusing on the uses of algorithms in relation to data surveillance, data retention and data mining. The picture they are presenting, however, is not as positive and optimistic.

In responding to Bennett's arguments on the uses of algorithms, I want to focus on one aspect that is missing from his article: data collection and, in particular, the ethics of data collection. To be able to benefit from the full potential of algorithms, we must have data, and lots of it. Data, and particularly 'big data', does not magically appear in our hard disks in searchable and machine-readable fashion. Relevant and useful data with which to study international politics is not naturally out there in the public domain and it cannot be retrieved simply by a Google search. Yes, the examples Bennett provides – the count of the frequency with which US newspapers mentioned the names of different Iraqi leaders and its relation to their perceived significance, and the counts of the frequency 'with which US and Chinese media use different formulations to define US and Chinese roles and relations'[2] – are based on publicly available texts. But the main potential of algorithmic processing, or automated reasoning, does not rest with processing a few hundred or few thousand documents. The 'real' potential of algorithms for social science research resides in their ability to search, process and relate millions and billions of documents and data markers. That's what algorithms do best. Algorithms can provide pretty accurate analysis of real-time world events; they allow us to map out relationalities of actors and actants. But they can only do that if we have the necessary data.

If we can provide a fuller picture of a particular political event, or if we can understand global trends in a more complete way, what will stop us? Should anything stop us? The potential of 'big data' analysis through algorithms rests on the scale of the data available to the researcher. If the data is not publicly available but can be made available through personal relations with the foreign policy or security and intelligence communities or elites, is that an ethical way of doing research? Building on these questions and responding to the main argument of the article, I want to ask Bennett if there should be any (ethical) limitations on the ways we collect data. What kinds of data collection methods are ethically admissible for the type of mixed method computer-assisted research that he is proposing?

I find the significant silences in regard to data collection in Bennett's text problematic. The collection of useful, 'searchable and machine-readable' data on a scale that makes the

2. Andrew Bennett, 'Found in Translation: Combining Discourse Analysis with Computer Assisted Content Analysis', *Millennium: Journal of International Studies* 43, no. 3 (2015): 993.

use of algorithms meaningful for the purposes Bennett is suggesting is a serious undertaking. To acquire data at that scale, we must mine data or have open access to a database that already provides us with the collected data, and, ideally, at a mass scale. Such a mining process will be by design indiscriminate of the consent of the 'unknowing' research subject. That would inevitably put the practices of the researcher and the privacy of the research subjects on a collision course. Most research universities, including Dr Bennett's own institution, would have ethics boards to review human subject research extensively and carefully. Human subject research must assess the impact of the research project on the safety, privacy and well-being of the research subject. The same level of attention to ethics, however, is not yet present in relation to the mining of personal data and the privacy implications of the uses of 'big data' in social sciences research.

The kind of data that is most commonly used by 'big data' researchers to bypass this problem is metadata. Metadata describes other data. The kind of metadata that will prove to be most useful for social sciences research is 'descriptive metadata'. This kind of metadata describes instances of application data and/or data content. A common argument used in defence of privacy violations that occur as a result of the use of descriptive metadata is that metadata is not personal. It is often claimed that descriptive metadata does not violate personal privacy simple because it contains nothing personal about the user. If the recent NSA leaks by Edward Snowden are any indication, algorithmic processing and automated reasoning can deduct a great deal of information from metadata alone. Metadata, as such, is personal, and processing metadata for research purposes does have privacy implications that we should be aware of.

These privacy concerns might not be an issue when we are researching press coverage of a particular event or official documents that are in the public domain, but they become a serious issue when we shift the focus of our research to tweets or other social media posts to gather information on public opinion or the role of informal networks in political mobilisation. This is the inevitable direction that such an 'algorithmic turn' might take. Social media posts do reveal a great deal of personal information about the individuals that post them. Tweets, for example, often include geographical data. By focusing on the individual tweeting, we can identify their social networks, their relationalities within that network. Furthermore, based on the frequency of certain 'keywords' in their posts, we can identify a great deal of other personal information about the individual.

Just as quantitative research and formal modelling produced 'actionable' academic research that proved to be very attractive for the policy elite, algorithmic research based on 'big data' will also prove attractive. For scholars of international politics, there is a great deal of pressure from universities to interact with policy elites and governments in order to be 'policy relevant' or have 'real world impact'. However, such a relationship in the absence of an ethical code of conduct will, by design, have risks associated with the privacy, rights and liberties of the unknowing research subjects. This is the often-ignored politics of algorithmic research methods. These methods represent an effective tool of research, but without an ethical code of conduct they also pose a great deal of risk to personal liberties, rights and privacy due to the mass-scale data mining necessary to use them to their full potential.

While I do believe that algorithmic data processing and automated reasoning will inevitably become more common in textual data processing, and perhaps might even be

the basis of cross-pollination of methods through increased use of mixed-methods research, prior to this proliferation we must reflect seriously on the ethics of this approach. Post-Snowden, a larger proportion of the general public has been exposed to the dangers of data mining, techniques used to gather mass-scale information, and the effects on privacy. Those that have been working on surveillance studies already knew about the dangers of surveillance for rights and liberties. Any discussion of algorithmic data processing without a discussion of ethics in relation to the data collection process comes across as either extremely naïve or deceptive.

In anticipation of this 'algorithmic turn' in social sciences methodologies, university ethics boards and professional organisations such as the International Studies Association (ISA) must form expert committees and develop a code of conduct that will guide researchers on best practice in their respective fields. These committees must be representative of the diversity of the discipline, in terms of gender, geographical origin, race and intellectual perspectives. Without a pre-emptive oversight mechanism that is embedded in the institutional architecture of the field – research methods courses, grant applications processes, tenure evaluations and ethics board considerations – algorithmic research methods have the risk of sharing the same fate as some of the earlier human-subject experiments. Algorithms and computer-assisted research methods might have the potential to bring different epistemological perspectives closer. But introducing these methods without proper ethical oversight comes with its own risks and challenges. The potential contributions of algorithmic data processing and automated reasoning must be evaluated and contextualised in relation to the ethical considerations. These are the stakes for algorithmic methods that Andrew Bennett fails to mention in his article.

Acknowledgements

The author would like to thank Sarah Mutlu and Cora Lacatus for their valuable comments and feedback.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author Biography

Can E. Mutlu is Assistant Professor of International Relations at the Bilkent University in Ankara, Turkey. His research interests intersect three areas, security, science and technology, and international political sociology with a specific focus on emerging research methods. He is the co-editor of *Critical Methods in Security Studies: An Introduction*. His recent research appears in *Comparative European Politics, European Journal of Social Theory, Eurasia Border Review, Environment and Planning D: Society and Space and the Review of International Studies, Critical Studies on Security* and *International Political Sociology*.