

Strengthening the evolutionary social sciences with more data, less ‘theory-worship’

Rebecca Sear, London School of Hygiene and Tropical Medicine

Comment on Barrett (2020) Deciding what to observe: Thoughts for a post-WEIRD generation, for special 10th anniversary of the WEIRD paper, edited by Coren Apicella, Joseph Henrich, Ara Norenzayan.

Published in Evolution & Human Behavior 2020

DOI: <https://doi.org/10.1016/j.evolhumbehav.2020.07.010>

The evolutionary social sciences have made significant progress over the past almost 50 years, developing from a niche and controversial endeavour to research which is now firmly embedded in at least some social science disciplines, notably psychology and anthropology. Clark Barrett’s excellent position piece takes stock of the successes of the evolutionary social sciences, but also highlights areas which need strengthening (Barrett, 2020). Here I propose that careful use of secondary data is one solution to a key weakness that Barrett highlights: the evolutionary social sciences do not pay sufficient attention to data.

A clear strength of the evolutionary approach to human behaviour is its strong theoretical framework; unfortunately, this focus on theory can become a weakness, because of what Barrett refers to as ‘theory-worship’. There has been a tendency in the evolutionary social sciences to pay too much attention to theory, to move too quickly to decide that particular hypotheses are indeed correct, without conducting sufficient empirical analysis to be confident of this conclusion.

A further problem with data in the evolutionary social sciences is the assumption that any old data will do to test hypotheses; an assumption which Barrett’s analysis shows is still sadly widespread. As a result, what evolutionary social scientists know about human behaviour is largely based on convenience samples of WEIRD students or MTurkers, with only a small secondary node of research in small-scale societies (typically described as those largely living a subsistence lifestyle with limited market integration). There is relatively little research on the vast swathe of humanity who are neither college students nor living in small-scale societies.

Improving respect for data in the evolutionary social scientists may be achieved by making greater use of secondary data (defined as data not collected by the researcher analysing the data). As researchers in the human sciences, we are fortunate in that we have access to very large amounts of existing data on our study species, and can draw on considerable expertise in social and health science disciplines, such as demography, sociology and epidemiology, in the analysis of such data. Developing knowledge and understanding of existing secondary data sources, how they are collected, and the skills needed to analyse such data, would broaden and deepen evolutionary social scientists’ ability to test empirical hypotheses, as well as having the enormously beneficial side effect of greater integration with other social and health sciences.

The use of secondary data is not yet widespread in the discipline, nor is training in accessing and analysing these sources very common in evolutionary social science programmes. This is despite their use in some pioneering work in the discipline, and a steady increase in their use, from a low base, in human behavioural ecology (see Stulp, Sear, & Barrett, 2016 for further discussion of the benefits of using secondary data in the evolutionary social sciences). Yet many different sources of secondary data exist, including both population-level and individual-level data, from countries across the world, often freely available to researchers (see [here](#) for a spreadsheet with links to secondary data resources).

Administrative data – data collected by governments through data collection exercises such as censuses – is one important source (Connelly, Playford, Gayle, & Dibben, 2016). Surveys are another source; these are very variable but may contain rich, individual-level demographic, socioeconomic, health, time-use and other data. Surveys may be available for a single country – often designed to be a representative sample of the national population, though some geographically-limited surveys are available. Other surveys are multi-country data collection exercises, where similar data is collected across several different countries, so that comparative analysis can be performed.

Other sources of secondary data exist, and have been used in the evolutionary social sciences, but which are not always publically accessible. These include historical data from church and civil records, which may involve the creation of databases from historical archives; and data from Health and Demographic Surveillance Sites (HDSSs), which are ongoing longitudinal data collection exercises in lower- and middle-income countries (Ye, Wamukoya, Ezeh, Emina, & Sankoh, 2012). Access to HDSS data typically requires developing collaborations, and funded research projects, with researchers based in HDSSs. A significant advantage of such sites is that the infrastructure available allows for the collection of new data, alongside the use of existing data.

Secondary data typically have the advantage of larger sample sizes than is possible through primary data collection, and some data collection exercises extend over many years. The most useful sources are longitudinal, meaning that data are repeatedly collected from the same individuals and/or area over time, which makes life course analyses possible. Longitudinal data, and large sample size, allow for sophisticated statistical modelling, including analysis of change over time. The breadth of data they contain allows for exploratory analysis, and for testing multiple hypotheses or explanations.

Secondary data analysis has its limitations. Researchers are restricted to the data contained in the surveys, and secondary datasets often sacrifice depth for breadth and sample size. Secondary data also need to be treated with caution, as they may contain errors and biases (as the recent furore over the use of a wholly inadequate dataset purportedly of ‘national IQ’ demonstrates: Ebbesen, n.d.). However, all methodologies in the human sciences have their limitations and, as long as researchers exercise due diligence before embarking on data analysis, use rigorous methods to analyse data, and interpret results with appropriate caution, the use of secondary data can become an extremely useful tool in the toolkit of the evolutionary social sciences.

The investment of time needed to develop the skills for data management, analysis and interpretation should not be under-estimated. Secondary datasets may be very large and complex, requiring substantial preparation before they are suitable for analysis; sophisticated statistical analysis to understand associations between variables; and rigorous documentation during preparation and analysis (see for tips on the use of secondary data: Rosinger & Ice, 2019; *Second. data Anal. An Introd. Psychol.*, 2010; Vandembroucke et al., 2007). But the development of such skills across the evolutionary social sciences will position the field to take advantage of the ‘big data’ revolution, which is gaining momentum in the social sciences.

Secondary data also raise issues around ‘researcher degrees of freedom’: the large amount of data available means that researchers have many different options about exactly how to test a particular hypothesis, which may result in difficulty replicating results, or ‘fishing’ exercises, where researchers trawl datasets for any ‘interesting’ findings, rather than conduct theoretically-motivated hypothesis tests. Pre-registration (van den Akker et al., n.d.) and ‘multiverse’ analysis (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016) may help to avoid these pitfalls.

This call for more emphasis on data in the evolutionary social sciences should not be read as a call to abandon theory. Data and theory are stronger together, and should be used in an iterative fashion; evolutionary researchers should not simply see data as a means of confirming their hypotheses, but should be more open to what data has to tell them, and to modifying their hypotheses based on data analysis.

The social science disciplines which have adopted the evolutionary approach most enthusiastically – psychology and anthropology – are those which tend to prioritise primary data collection, and may be somewhat sceptical about the use of secondary data. But careful use of secondary data ought to become a more prominent methodology in the evolutionary social sciences: this will strengthen the ‘observational’ basis of the discipline, allowing for more thorough description of human behaviour, so we have a better idea of what we need to explain with our theory; it will also broaden the samples used in our field.

References

- Barrett, H. C. (2020). Deciding what to observe: Thoughts for a post-WEIRD generation. *Evolution & Human Behavior*.
- Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, 1–12. <https://doi.org/10.1016/j.ssresearch.2016.04.015>
- Ebbesen, C. L. (n.d.). Flawed estimates of cognitive ability in Clark et al. *Psychological Science*, 2020. Retrieved June 21, 2020, from <https://psyarxiv.com/tzr8c>
- Rosinger, A. Y., & Ice, G. (2019). Secondary data analysis to answer questions in human biology. *American Journal of Human Biology*, 31(3), e23232. <https://doi.org/10.1002/ajhb.23232>
- Secondary data analysis: An introduction for psychologists*. (2010). *Secondary data analysis: An introduction for psychologists*. American Psychological Association. <https://doi.org/10.1037/12350-000>
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, 11(5), 702–712. <https://doi.org/10.1177/1745691616658637>
- Stulp, G., Sear, R., & Barrett, L. (2016). The reproductive ecology of industrial societies, part I: Why Measuring Fertility Matters. *Human Nature*, 27(4), 422–444. <https://doi.org/10.1007/s12110-016-9269-4>
- van den Akker, O. R., Weston, S. J., Campbell, L., Chopik, W. J., Damian, R. I., Davis-Kean, P. E., ... Bakker, M. (n.d.). Preregistration of secondary data analysis: A template and tutorial. Retrieved June 21, 2020, from <https://psyarxiv.com/hvfmr/>
- Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., ... Egger, M. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Medicine*, 4(10), e297. <https://doi.org/10.1371/journal.pmed.0040297>
- Ye, Y., Wamukoya, M., Ezeh, A., Emina, J. B. O., & Sankoh, O. (2012). Health and demographic surveillance systems: A step towards full civil registration and vital statistics system in sub-Saharan Africa? *BMC Public Health*, 12(1), 1–11. <https://doi.org/10.1186/1471-2458-12-741>