Methods for using qualitative data to inform behavioural rules in Agent-Based Modelling: Preliminary results of a Systematic Literature Review

Author1[1], Author2[2], Author3[3]

[1] Affiliation1
[2] Affiliation 2
[3] Affiliation 3

Abstract. Qualitative evidence (e.g., texts gained from transcribing interviews with or observations of people) is a suitable source for giving clues as to the in-context behaviour of individuals. As such, it is a natural source for informing the specification of corresponding agent behaviour within simulations. However, *how* to use this evidence to inform these behavioural rules is not straight forward. This study uses a Systematic Literature Review to identify such methods. Here we present some of the preliminary results of this review.

Keywords: Agent-Based Modelling, Methods, Qualitative Data, Systematic Literature Review.

1   Introduction

Qualitative evidence (e.g., texts gained from transcribing interviews with or observations of people) is a suitable source for giving clues as to the in-context behaviour of individuals [1]. As such, it is a natural source for informing the specification of corresponding agent behaviour within Agent-Based Models (ABMs). However, *how* to use this evidence to inform these behavioural rules is not straight forward. This study uses a Systematic Literature Review to identify such methods. Here we present some of the preliminary results of this review, in order to raise awareness of the topic among the participants at the Social Simulation Conference 2023 and to discuss these results with the community.

2   Methodology

A Systematic Literature Review (SLR) was employed to answer the research question. An SLR is a structured, transparent, and reproducible process to identify, select, critically appraise, and synthesize all of the documents meeting pre-specified eligibility criteria. In this study, we have used the SLR phases described in [2].

Search strategy. A first search was performed in September 2020 and a second one in June 2022 (to supplement the dataset with items published in between the two dates).

2

Seven databases were used: ACM Digital Library, IEEE Xplore, JSTOR, ProQuest, Science Direct, Scopus, and SpringerLink. In addition, two journals were searched, using their own website: Journal of Artificial Societies and Social Simulation and Journal of Simulation. The search was limited only to items published in the English language. No other limitations were used. The only inclusion criterion applied was that the paper should include a detailed description of how qualitative data was used in an ABM.

Based on this search strategy, this search query was used:

TITLE-ABS-KEY ("qualitative data" AND "agent-based")

Screening and analysis. The items identified were divided equally among the three co-authors, who then performed a first, broad screening of titles, abstracts, and keywords using the inclusion criterion. A second, strict screening of the retained items was performed on full-texts and any annexes/supplementary material, by the same person as in the 1st screening. The same procedure was followed for both search waves. All items retained for in-depth analysis and synthesis were equally divided among the three co-authors and coded following a coding scheme that included questions about the paper itself (e.g., theoretical, methodological, applied, mixed; focus on the method of using qualitative data in ABM, focus on the ABM, both), the qualitative data (e.g., data source, purpose of data, type of data, how was data collected and processed), the ABM and modelling process (e.g., model purpose, application domain, use of documentation framework), and about the qualitative data to ABM mapping (e.g., degree of mapping detail, what was the data used for during modelling, what mapping approach was used). At the beginning of the coding process, the three co-authors independently coded the same three items and, during several discussion sessions, harmonised their understanding of how the coding should be performed. All items that raised questions during the individual coding were discussed among all the co-authors.

3   Results and discussion

3.1   Metadata

The two search waves returned 1884 items in total. Only 159 'survived' the first screening and of these, 74 were retained for in-depth analysis and synthesis. Figure 1 depicts the number of papers undergoing the second screening (labelled as "all") and the number of retained papers (labelled as "relevant") papers by year.

Most of the relevant papers are published in the last decade; the establishment in 2013 of the ESSA Special Interest Group on the topic of using qualitative data to inform behavioural rules in ABMs being most probably a driver for this. Most of the items (49; 70%) are published as journal articles; 19 items (27%), in conference proceedings; five as book chapters; and one as a thesis. As publication venues, not surprisingly, the Journal of Artificial Societies and Social Simulation (JASSS) is the main outlet (16; 22%), followed by the proceedings of the Social Simulation Conference (SSC) (8; 11%), the

Journal of Environmental Management (4; 5%), and Environmental Modelling & Software (3; 4%). Surprisingly, none of the journals is an outlet specifically connected to methodology aspects of social sciences. The fact that almost all but two of the proceedings are connected to computer science and simulation is something most likely due to the culture of publishing such proceedings in these disciplines. JASSS and SSC are the main outlets for publishing these studies (24; 33%). The rest of the 50 papers are scattered among 33 other publication outlets.

Moreover, also not surprisingly, almost 50% of the papers (35) have two or three authors; around 30% (23) have four to five authors; around 11% (8) have one single author, with the same number of papers having more than six authors (with 25 being the highest number of authors). Thus, the vast majority of papers is written by small (2-3) to medium-sized teams (3-5), possibly pointing to the scale of the corresponding effort when using qualitative data to inform ABMs.
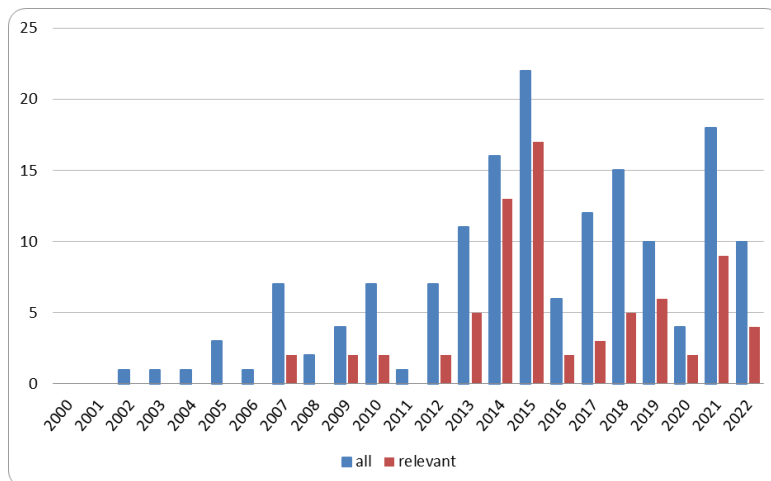

Figure 1. Number of relevant papers and total number of papers, by year.

3.2 Codebook data

Out of the 74 items retained for in-depth analysis and synthesis, about 30% (22) include a more or less detailed description of how the qualitative data was used when informing the ABM. Considering that the expectation of methodological rigour and transparency of modelling applies also when using qualitative data, this is a rather small number.

Out of the 49 journal papers, most (17) focus on the model (i.e., the model of the scenario relevant to respond to the research question of concern), followed by a focus on the method of mapping qualitative data into the model (14; 29%; most in JASSS), both model and method (12), or something else (6). Out of the 19 conference proceedings papers, the majority (10; 53%; most in SSC) focuses on the method, followed by a focus on something else (4), on both model and method (3), or on the model (2).

Most of the papers (31; 41%) did not include any documentation framework (e.g., ODD protocol).

Several methods were used for collecting the qualitative data used in the ABMs included in this study, with a mix of methods being the most popular one (32; 43%), followed, at a big distance, by individual interviews (16; 22%), and surveys (4; 5%). In the case of four papers (5%), this data collection method was unclear. Conventional literature review, participatory process, participatory observation, and focus group were also represented in the dataset, but in two or less papers per each of these methods. Within the mix of methods, individual interviews are the most popular (80%), followed by literature/documents (35%), other methods (28%), observation and survey (each 25%), focus group (19%), and participatory process (10%). Almost 40% of the papers used secondary data; 35% used primary data; and 23% used a mix of both. In almost 50% of the papers, the qualitative data was collected for the purpose of the ABM, in 20% of the papers this is not clear. In almost 80% of the cases, the data was not made publicly available.

With respect to the application domain of the ABM, more than 40% of these were applied in socio-ecological systems / environment-related cases (e.g., fishing, eco-tourism, viticulture, bushmeat hunting, land-use). Social psychology (e.g., emergency behaviour) and innovation studies follow at a big distance (around 9% each), and 12 other domains share each in between 2% and 7% of the studies included in this SLR.

With respect to the methods of mapping qualitative data into the ABM, we have identified nine very well described individual methods. We are now analysing these together with the rest of the data, trying to find common way(s) to describe them.

4 Preliminary conclusions and further questions

These are our preliminary conclusions and questions for the community:

- Few papers include a detailed enough description of how qualitative data was used when building the ABM. Is this practice a potential indicator of lacking methodological rigour in agent-based modelling?
- Conference proceedings seem to be the venue where such detailed methods are published. Why not in methodology-oriented journals, so that they are easily found by new practitioners?
- JASSS and SSC are the main outlets for publishing these methods (24; 33%). The rest of the 50 publications are scattered among 33 other outlets. Are we preaching to our own choir?
- Most of the items were categorized within the socio-ecological systems domain. Are there no other domains relevant for ABMs and qualitative data?
- There is no widely used method to map qualitative data into an ABM. Most of the well explained methods are used sparsely and sometimes only by the person(s) who described them initially. Why are these methods not used by more people?
- There is no harmonised or standardised way to describe this mapping yet. Is it possible to find one (or some), and if so, would it be useful?

References

1. Antosz, P., Bharwani, S., Borit, M. and Edmonds, B., 2022. An introduction to the themed section on 'Using agent-based simulation for integrating qualitative and quantitative evidence'. International Journal of Social Research Methodology, 25(4), pp.511-515.
2. Bouck Z, Straus SE, Tricco AC. Systematic Versus Rapid Versus Scoping Reviews. Methods Mol Biol. 2022;2345:103-119. doi: 10.1007/978-1-0716-1566-9_6. PMID: 34550586.