Challenges of designing and implementing simulation models of peer review

Thomas Feliciani a*, Pablo Lucas a*, Junwen Luo b, Kalpana Shankar b

a School of Sociology and Geary Institute for Public Policy
b School of Information and Communication studies and Geary Institute for Public Policy
University College Dublin, Ireland
* corresponding authors: {thomas.feliciani, pablo.lucas}@ucd.ie

Abstract

Science relies on peer review. Through this mechanism, manuscripts are selected for publication and grant proposals for funding. However, the processes of peer review do not operate in a vacuum; they reflect the priorities, norms, and practices of the institutions in which they are embedded, such as scientific communities, funding agencies, publishers, and scholarly societies, each with their own perspectives and logics (Bollen et al. 2014; Benner & Sandstrom 2000). Peer review is a multi-level system. At the macro level a funding agency sets its priorities and goals for funding based on national priorities and legal mandates. At the meso level, funding agencies use peer review to select which proposals to fund, but also integrate their own strategic objectives (gender balance, geographical diversity, disciplinary needs for example) into the selection process. At the micro level, individual reviewers and panels bring their own perspectives to bear on the review processes. In particular, the dynamics of meso- and micro-level complexity provides an area of exploration that could benefit from simulation studies for two reasons. Simulation studies help us understand what features of the peer review process emerge from different norms, relationships, attitudes and behaviors of the actors and organizations involved. These methods also allow us to develop
and test policy recommendations for the improvement of peer review in these same organizations.

In our own project we started by mapping existing simulation models of peer review and identified knowledge gaps in the literature, then started developing a simulation model to address these gaps. We found that numerous researchers had studied peer review systems by means of formal and computational modeling, such as agent-based models (ABM) (Squazzoni & Takács 2011). We counted 44 papers on simulation models of peer review published since 1969: some were used to compare the efficiencies of alternative peer review systems (e.g. Kovanis et al. 2017); some compared different behavioral strategies of authors, editors or reviewers (e.g. Thurner & Hanel 2011; Squazzoni & Gandelli 2013); some sought the origin of the issues of peer review, such as biases, high costs and inefficiencies (e.g. Righi & Takács 2017).

Reviewing existing models

To perform a scoping review of the literature (in process), we ran queries on Scopus and Web of Science to find a comprehensive set of publications on simulation models of peer review. We integrated the set by reference chaining and with papers from our knowledge. We then classified the models based on the kind of models, the kind of peer review systems, the prominent model features, and the research questions explored with the model. Besides proposing a taxonomy of models of peer review, our scoping review identifies some open issues and knowledge gaps.
First issue: limited model integration

In our review we found a highly fragmented landscape of models, assumptions, and findings. None of the papers we reviewed attempted to compare previous models, and only in a few cases was a model further developed after its initial publication. The lack of integration between models carries important consequences for the generalizability of their findings. This is the case for some key model assumptions, which constitute the difference between some models. An example is the assumption that submissions (e.g. manuscripts or grant proposals) have intrinsic, “true” quality. Whereas some models are built on this assumption, other models assume no intrinsic quality of these submissions. The intrinsic quality assumption represents a modeler’s perspective about the role of reviewers: if submissions do have an intrinsic quality, then it is the reviewer’s task to estimate the quality as accurately as possible. Conversely, assuming no intrinsic quality implies that reviews are purely subjective, and disagreement between reviewers does not imply that some reviewers are wrong. We ask: would a model’s predictions be different if we did (or did not) assume that submissions have an intrinsic quality? As models with these two alternative modeling assumptions have never been systematically compared, we do not know the extent to which their findings are robust to the assumption.

Second issue: lack of empirical data

Despite growing interest and calls by the computational modeling community for the empirical calibration and validation of simulation models (Hedström & Manzo 2015), we found that only a minority of papers made use of empirical data. Access to peer review data is difficult, as it does require an appropriate management protocol to ensure both confidentiality and anonymity. When empirical data was used, few model parameters were
calibrated based on that, and few model predictions were compared to the available empirical data.

**Work in progress**

We are developing an ABM of peer review process at one national funding agency that: (1) integrates features from existing relevant models in literature, and (2) is empirically calibrated and validated with qualitative and quantitative data, including textual data from policy and organizational documents and interviews.

The integration of features from previous models (1) is done by ‘aligning’ alternative implementations found in the literature (see e.g. Axtell et al. 1996). This implies that in the simulation environment of our ABM we are able to compare alternative assumptions/features from previous work, all other factors being kept constant. This allows us to test the effects of these assumptions and the robustness of our findings against them.

The use of diverse data sources for ABM calibration and validation (2) will be done by integrating insights from both qualitative and quantitative data sources. This presents some methodological challenges. On the one hand, it is clear that expert interviews can be useful to inform a simulation model on the functioning of the peer review process. On the other hand, there are no guidelines or best practices for the task translating interview transcripts into a formal system. Relatedly, it is not clear how to handle discrepancies between different data sources, e.g. when micro- and meso-level actors provide contradictory descriptions of peer review practices, or when their description clashes with the quantitative evidence. With our ongoing modeling work, we are currently considering ways to overcome these challenges.
References


