Short Writing Assignment: Safety in Numbers

Tom Wallace

Spring 2018

Research Question

Like most papers we have read in this course, this one does not have a research question. I would characterize the paper as more of an engineering document than an scientific publication. This is not a criticism: it describes a useful technical solution to an important real-world problem. But it does not develop or test falsifiable hypotheses, the Popperian *sine qua non* of science.

Model Effectiveness

This model is intellectually interesting and well-designed, but judging its effectiveness is not possible by any rigorous criteria.

The model does many things well. It offers a convincing theoretical exploration of the non-linearity and structural instability of crowd behavior, which justifies the individual agent-oriented nature of the model. It builds a simple structure—essentially, random walk with ant colony-style swarming behavior—that nevertheless produces interesting emergent behavior and macro-structure. It is well-documented, offering both programming-style flowcharts of model logic and detailed mathematical description of behavior. The model is well-designed as a policy tool, as it explicitly allows for the introduction of policy interventions—e.g., different configuration of entry and exit points, crowd control set-ups, and the like—and observing the effect on crowd flows. The paper's references to various planning committees and decisions suggests that the model was found believable and useful by non-technical decisionmakers, which is no small feat. Also, the early publication date of this paper (2002) makes all of the above even more impressive.

However, the reader has no intellectual basis for ascertaining whether the model actually works or not. The authors mention competing theories and models of crowd behavior, but do not attempt to compare this model's performance to them by some common measure. Is this model better or worse than Borgers and Timmerman (1986)? How about Kirchner and Schadsneider (2002)? We do not

know. The authors also did not validate their model against real-world data. They did use data to tune input parameters, but then did not take their tuned model and apply it some other dataset. This distinction between training and testing sets is critical. It is easy to over-fit models to one idiosyncratic dataset, when the goal is to capture the true nature of some broader phenomenon. I do not necessarily fault the authors for these choices because the motivation for their model is engineering good policy, not conducting science, but it does mean that the reader has no real basis for judging the model's performance.

Course Themes

While not a formal course theme, I think the following merits extended discussion: who is using agent-based models to do meat and potatoes science? We have conducted a wide-ranging literature review of spatial agent-based modeling, and I can count on one hand the number of papers that would meet the classifical definition of competing hypotheses compared and tested by their ability to explain empirical phenomena. Some models merely assess qualitative correspondence to theory (which is fine - this is a useful activity for theory development); others, such as this model, fit their model to empirical data but do not compare that performance to competing models (again, fine as a first-order sanity check); very few have taken agent-based model A and stacked it up against agent-based model B in terms of performance on some common empirical dataset. To continue the food metaphor, I would characterize the former two as "appetizers" and the latter as the "main course": some theory-building and calibration is well and good, but really exists in order to enable the comparison of competing hypotheses, and this final step is mostly absent in the ABM literature. I believe the root cause of this lies in the "one-off" nature of many ABMs, which I have mentioned before. For whatever reason—the still-incipient nature of ABM; inter-disciplinary boundaries; professional and publication incentives—most authors do not compare the performance of their model to that of other models (ABM or not). This is something that we should aspire to change as emerging scholars in the field.