# Trapping Explanations and External Shocks

November 15, 2018

## 1 Trapping Explanations

Many agent-based models in social epistemology and philosophy of science are used to construct *trapping* explanations; they provide explanations for the evolution of an epistemic community towards a steady state, which either with certainty or probability close to one "traps" the community, i.e. confines it to this state. Two well-known examples are the epistemic landscape model (Weisberg and Muldoon, 2009) and the bounded confidence model of opinion dynamics (Hegselmann et al., 2002).

In the former model, a set of researchers explores an epistemic landscape, eventually converging to the peaks of this landscape, from which the modeled agents never or almost never – depending on the exact agent type – move away; they become trapped. This result can be utilized to make a normative argument about the advantages of agent diversity, or as an explanation for a communities convergence to a suboptimal local maximum of inquiry.

The bounded confidence model represents a set of agents, updating their beliefs in light of all the other agents' beliefs. They employ equal weight linear averaging, but only across the subset of agents whose beliefs fall into their range of confidence. Once again, the model converges to a stable steady state. The exact state, depending on initial distribution and confidence range, may contain a single or multiple different opinions camps the agent set converged on. Steady states of the bounded confidence model with a small number of resulting opinion camps are utilized in explaining opinion polarization.

These are not the only instances of trapping: The epistemic network learning model by (Zollman, 2007), the social collaboration model by (Hong and Page, 2004) or Bicchieri's model of norm formation (Bicchieri, 2005, ch. 6) and various models built on the basis of these models such as (Holman and Bruner, 2015) or (Hegselmann and Krause, 2015) provide further instances.

This type of model is often used to explain the prevalence of a state in the target system represented by the trapping state: The bounded confidence model, for example, is supposed to explain polarization; Zollman's model is offered as an explanation of the survival of outdated medical theories. This explanatory power has been challenged on various grounds (Alexander et al., 2015; Frey and Šešelja, 2018), but we want to discuss a different issue pertaining to the explanatory power of trapping states quite generally.

## 2 The Explanatory Power of Trapping States

The problem we want to address might be best understood with an analogy to historical explanations. Consider various suggested explanations for the fall of Rome: Economic crisis, barbarian invasion, internal political struggles, and so forth. All of these plausibly may have contributed to the eventual fall of Rome; however, none of these phenomena was fundamentally new to the empire. So if a certain factor is cited as an explanation, but that same factor was prevalent before, it cannot constitute a sufficient explanation.[1]

The analogy in dynamical models of social epistemology is that they lack a component that would eventually shake the target system out of the trapping state; such external shocks are absent, and hence the explanations generated from these models are inherently insufficient. To provide complete explanations, they not only had to represent the mechanism of convergence, but also the complementary force of epistemic shocks.

Before turning to a case study, we want to anticipate two possible objections: First, that the relevant difference is just the quantity of whatever defines the size of the external shock. This might very well be true, but to establish this explanation, and to precisely state what the size of a shock means, shock modules need to be added the models still. Second, it might be proposed that the models are not supposed to explain the eventual escape[2] from the trapping state. But any such explanation seems to be insufficient in a similar way as in the above example of explaining the fall of Rome: The mechanism represented in the model may contribute to the phenomenon in question, but even if it is part of an actual explanation, it remains necessarily partial until it accounts for potential events of the type that eventually moved the system away from its steady state.

[1]cf. Martin (1979) for a general discussion of related problems in historical explanations.

[2]Note, that while the language may suggest a trapping state to always be negative, and many of the models are interested in explaining the collective convergence to epistemically undesirable states; in general, however, a trapping state can just as well be positive, that is, epistemically or otherwise desirable.

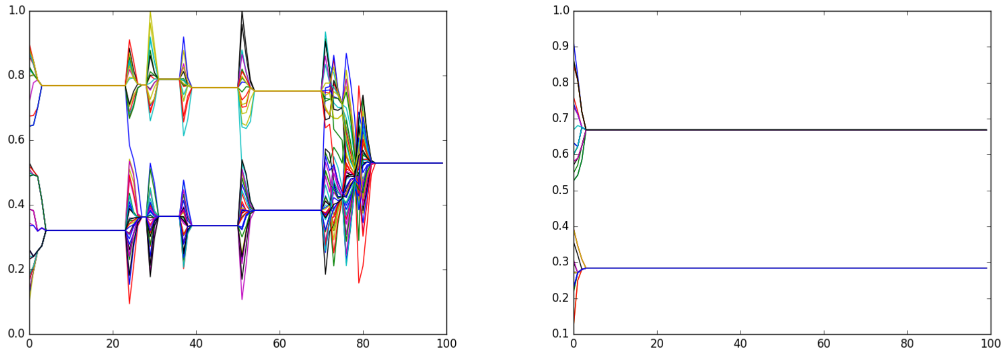


Figure 1: Bounded confidence model with (left) and without (right) external shocks, example run. Parameters: $n = 50$, $\mu = 0.1$, $\sigma = 0.1$, $r = 0.2$.

# 3 Modeling External Shocks

The bounded confidence model is a structurally simple instance of a trapping model, and can easily be extended with an external shock model. The model is described by

$$o_i(t+1) = \frac{1}{|P_i|} \sum_{j \in P_i} o_j \tag{1}$$

where $P_i$ is the group within $i$'s two-sided confidence interval.

We extend it with random global displacement events. At each time step, with probability $p$ every agent is randomly displaced by $\epsilon_i$, which is drawn from a normal distribution with mean $\mu$ and standard deviation $\sigma$.[3] To remain within model parameters, the resulting opinions are cut off at zero and one. An example run is depicted in Figure 3. The resulting numbers of opinion camps are depicted in figure 3 for the same parameter configuration.

This is not the only possible shock model for the bounded confidence model, and we have not yet started to explore in detail the manifold options to introduce shocks into the models mentioned in the beginning. One problem well known to agent-based modelers is that additional modules or subprocesses increase the difficulty to fully analyze the model. Hence, we suggest to introduce such models at least at first in the simplest member of each model family, and start exploration with specific questions.

As a concluding remark, we want to point out that introducing external shocks is not meant to increase realism – which may happen as a side effect – but to enable them to provide more complete explanations, regardless of

[3]To clarify: All agents are displaced synchronously, but with independently drawn $\epsilon$.

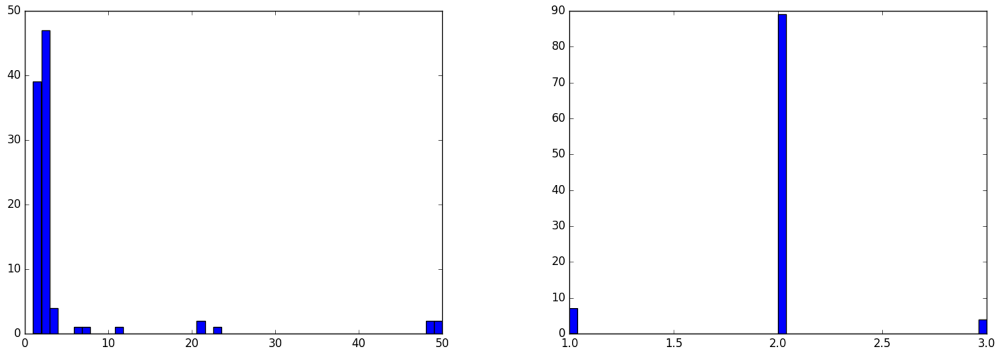


Figure 2: Number of opinion camps after 100 timesteps with (left) and without (right) shocks. Parameters as before.

their status as potential or actual.

# References

Alexander, J. M., J. Himmelreich, and C. Thompson (2015). Epistemic landscapes, optimal search, and the division of cognitive labor. *Philosophy of Science 82*(3), 424–453.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Frey, D. and D. Šešelja (2018). Robustness and idealizations in agent-based models of scientific interaction. *The British Journal for the Philosophy of Science*.

Hegselmann, R. and U. Krause (2015). Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model. *NHM 10*(3), 477–509.

Hegselmann, R., U. Krause, et al. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of artificial societies and social simulation 5*(3).

Holman, B. and J. P. Bruner (2015). The problem of intransigently biased agents. *Philosophy of Science 82*(5), 956–968.

Hong, L. and S. E. Page (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences 101*(46), 16385–16389.

Martin, R. (1979). Historical counterexamples and sufficient cause. *Mind 88*(349), 59–73.

Weisberg, M. and R. Muldoon (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of science 76*(2), 225–252.

Zollman, K. J. (2007). The communication structure of epistemic communities. *Philosophy of science 74*(5), 574–587.