Simulating Science: a Multiagent Model of Scientific Evolution

Institutionalized science is a social process. It can be construed as a collective cognitive system distributing knowledge production and evaluation between multiple agents. Taken together, these agents can be seen as implementing a search algorithm trying to maximize empirical adequacy in some theoretical space. There is a family of views in the philosophy of science—falling under the label 'Evolutionary Epistemology'—that would describe this algorithm as a 'genetic' one, where some or all of the constitutive objects of science (ideas, concepts, theories, etc.) are produced, transmitted and selected in a Darwinian fashion. The most ambitious and complete proposal along these lines is arguably David Hull's *Science as a Process* (1988). While Hull's theory is much detailed, it remains a verbal proposal that could benefit from formalization.

The project presented here models Hull's theory of the scientific process as a multiagent system and works out its consequences using computer simulation. My main argument tries to show that multiagent systems should become an important tool for philosophical work aiming to understand the scientific process and to establish epistemological norms for individual scientists, as well as for the social organisation of science. The argument does not rest on the validity of Hull's theory, or even on the adequacy of its formalisation. Its form is that of a proof of concept: given one characterization of the scientific process, here is how you can formalize it as a multiagent model and use it to produce norms. Also, this method is not meant to replace traditional epistemological tools, but to complement them. My whole talk will: 1) present Hull's theory; 2) introduce multiagent systems; 3) present a multiagent model of Hull's theory, and 4) show how simulations using this model can lead to the emergence of norms. Space constraints will however reduce this, here, to a very sketchy summary.

According to Hull, Darwinian evolution is a general process. It requires *interactors* and *replicators*. As long as the actions of the interactors cause the differential reproduction of the replicators, you have evolution. Organisms and their genes implement an instance of that process, but so do scientists and their ideas. In order to transmit their ideas, scientists must acquire *credit*, i.e. get to be respected by their peers, with whom they are both competing and collaborating. Citation of someone's work in a paper is a mechanism that trades credit for support. Sharing credit with your students provides something akin to "inclusive fitness", since they are vectors for your own ideas.

In a multiagent model, agents are usually heterogeneous. They can be more or less creative, communicative, sociable, aggressive, generous, etc. These personal characteristic can be expressed as real numbers that will be taken into account by the formal rules of behaviour implemented by the agents. Agents are also situated in an environment where interactions are mostly local. In our model, this environment is a social network: scientists interact with their students and their collaborators. They also write peer-reviewed articles that allow them to share their ideas with the whole community. These ideas are strings of information that can be formalized as vectors in a multidimensional space. When ideas are transmitted from one scientist to another, some bits of information can be mutated, introducing noise in the transmission. Since it is scientific ideas we are interested in, each idea is assigned some "empirical adequacy" value, using an arbitrary objective function. This function constitutes the problem space that the agents are working in. They do not have direct access to its values, but they can try to approximate them using more or less adequate "tests". Overall, we can measure the efficiency of the system by how close the subjective ratings given by the agents to their own ideas are to their "real", objective values.

The complexities of this model would make it analytically intractable, but computer simulation allow us to examine, one step after another, what happens in the system, and to see the effect that different parameters (like different distributions of personal characteristics) have on its efficiency at maximizing empirical adequacy. It is conceivable, for example, that too much interaction between scientists would lead the system to settle at a local optimum, while having small, close-knitted but independent communities would allow a more thorough exploration of the problem space. Different results here would suggest different norms as to how science should be socially organised.

Hull, D. L. (1988). *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.