Six challenges for fully autonomous scientific discovery

Benjamin C. Jantzen, Dept. of Philosophy, Virginia Tech

Science is a house of many mansions, a diverse collection of cognitive aims that include prediction, control, and understanding. It is famously iterative in the sense that the cyclical satisfaction of cognitive aims is necessary for advancement as, for instance, in the idealized hypothesize and test cycle of theory selection. It is also famously open-ended in that there is no clear end to the production of new scientific concepts and theories. Despite longstanding philosophical dismissal of the prospect, recently realized computational programs confront us with the possibility of capturing both of these aspects of science by automating not just particular methods of satisfying individual cognitive aims, but of chaining such methods together in complete iterative cycles. Whether it is possible for automated discovery to be open-ended in the same way as human discovery is unclear. But whatever the case, imagining a fully autonomous open-ended scientific discovery process is a valuable exercise in teasing out those aspects of scientific investigation that are yet to be precisely characterized. This in turn supports the development of tools to assist the scientist in carrying out those tasks. More radically, an explicit approach to completing an entire discovery cycle permits one to identify key steps that have remained invisible under our current accounting of the process.

Historically, the presumption of philosophers and computer scientists has been that the central objects of any discovery cycle are theories—explicit representations of claims about the contents of the world and their interactions. The introduction of novel scientific concepts appears to present an insurmountable obstacle to automation in this view because the origin of appropriate concepts is presumed by many to be an unanalyzable leap of creativity. However, recent work in the automated identification of novel variables using the "dynamical symmetries" framework suggests this is not where the obstacles to automation truly lie. To illuminate why this should be the case and where the genuine obstacles to automation remain, I adopt a perspective that foregrounds not theories but rather "natural kinds". These are classes of physical system—types of phenomena in the world—well-described by a shared set of natural "laws" or governing equations. Members of a natural kind are united by shared aspects of behavior (e.g., all radioactive decay is exponential) but generally vary in their individual behavior (e.g., some radioactive substances decay faster than others). What we generally call scientific theories are accounts of the shared behavior of members of a natural kind. With a focus on natural kinds, the central discovery cycle is that of identifying variation within a natural kind, introducing novel variables to account for that variation, then identifying new natural kinds in terms of those variables in which further variation can be noted.

I adopt this vantage point in order to dissolve some apparent challenges to fully autonomous discovery and to identify six impediments that are either unrecognized, poorly characterized, or addressable by methods that may be unfamiliar to the broader community interested in integrated computational discovery. Specifically, I consider the problems of: (1) identifying the qualitative contexts connected with behavioral variation within a natural kind, (2) constructing an appropriate quantitative scale of measurement for newly identified variables, (3) efficiently constructing an explicit theory of behavior within a natural kind, (4) forming a new "affordance", or in other words, constructing a new sort of experimental manipulation (5) productively guiding exploratory experimentation with a newly acquired variable, and (6) translating between potentially cross-cutting ontologies. For each challenge, I render the problem as precise as possible in terms of the inputs available from other pieces of the cycle, the necessary outputs to feed into downstream tasks, and the nature of the problem to be solved in between.

References representative of my work on computational discovery

[1] C. Shea-Blymyer, S. Roy, and B. Jantzen, *Entropy*, vol. 23, no. 9, p. 1191, Sep. 2021. [Online]. Available: https://www.mdpi.com/1099-4300/23/9/1191.

[2] B. C. Jantzen, *Philosophies*, vol. 6, no. 4, p. 103, Dec. 2021. [Online]. Available: https://www.mdpi.com/2409-9287/6/4/103.