# Science with Artificially Intelligent Agents: The Case of Gerrymandered Hypotheses

Ioannis Votsis
(University of Duesseldorf)
votsis@phil.uni-duesseldorf.de

Barring some civilisation-ending natural or man-made catastrophe, future scientists will likely incorporate fully fledged artificially intelligent agents in their ranks. Their tasks will include the conjecturing, extending and testing of hypotheses. At present human scientists have a number of methods to help them carry out those tasks. These range from the well-articulated, formal and unexceptional rules to the semi-articulated rules-of-thumb and intuitive hunches. If we are to hand over at least some of the aforementioned tasks to artificially intelligent agents, we need to find ways to make explicit and ultimately formal, not to mention computable, the more obscure of the methods that scientists currently employ with some measure of success in their inquiries. The focus of this talk is a problem for which the available solutions are at best semi-articulated and far from perfect. It concerns the question of how to conjecture new hypotheses or extend existing ones such that they do not save phenomena in gerrymandered or ad hoc ways. This talk puts forward a fully articulated formal solution to this problem by specifying what it is about the internal constitution of the content of a hypothesis that makes it gerrymandered or ad hoc. In doing so, it helps prepare the ground for the delegation of a full gamut of investigative duties to the artificially intelligent scientists of the future.

Perhaps the most famous example of a gerrymandered system of hypotheses is Ptolemaic astronomy. The Ptolemaic system was comprised of the geocentric hypothesis as well as several other auxiliary hypotheses, e.g. hypotheses that specified the radii of deferents and epicycles. As is well-known, astronomers working under this tradition kept tweaking the auxiliaries to adequately fit the known phenomena. At no point, had they considered tweaking or dropping the geocentric hypothesis itself, which, as we now know, is incorrect. Despite their short-sightedness, they were not acting so different to modern scientists. It is not easy to tell in advance whether attempts to fit the known phenomena will result in a gerrymandered system or one that adequately captures the natural order of things. Moreover, it is not just a question of identifying those systems or hypotheses that are gerrymandered as those that contain at least one false hypothesis or hypothesis-part respectively. After all, a system or hypothesis whose constituents are all true may still be gerrymandered or ad hocly put together. Finally, it is not, as some have suggested (see, for example, Lakatos 1968), a matter of being unable to predict novel phenomena with gerrymandered hypotheses, for, in principle, nothing prevents some such hypotheses from being able to make novel predictions.

This talk proposes an altogether different approach to the problem, one that focuses on the way that potential support for a system or a hypothesis is propagated through its content. Let us call a hypothesis or system of hypotheses 'gerrymandered' or 'ad hoc' if and only if some of its content parts are disjointed. Any two content parts expressed as propositions $A$, $B$ are disjointed if and only if $P(\alpha/\beta) = P(\alpha)$ for all propositions $\alpha, \beta$ where $\alpha$ is a relevant deductive consequence of $A$ and $\beta$ is a relevant deductive consequence of $B$. The first thing to note about the concept of disjointedness is that it is articulated in terms of the concept of *probabilistic independence*. The concept of probabilistic independence is apt here because it allows us to express the idea that two propositions are confirmationally unrelated. After all, the probability of the one is not affected if we assume the truth (or falsity) of the other. The second thing to note is that to establish the confirmational unrelated-ness between two propositions $A$, $B$ it is not enough to merely focus on the propositions themselves but must also take into account their *deductive consequences*. The reason for this is that two propositions may be probabilistically independent even though some of their deductive consequences are probabilistically dependent. To rule out such cases we must demand that probabilistic independence holds all the way down, that is, between all – save for an exception to be discussed below – the deductive consequences of two propositions. This demand is an apt way to express the idea that no part of the content of the one proposition confirmationally affects any part of the content of the other proposition. The third and final thing to note is that unless we restrict our evaluation to *relevant* deductive consequences of propositions the concept of disjointedness would be unsatisfiable. The idea of a relevant deductive consequence is fully developed in Schurz (1991): "the conclusion of a given deduction is irrelevant iff the conclusion contains a component [i.e. a formula] which may be replaced by any other formula, salva validitate of the deduction" (pp. 400-1). Here's why we need it. Whatever the content of propositions *A*, *B* we can always validly derive consequences that are common to both, e.g. $A \vee B$. The existence of such trivial common consequences guarantees that there is a pair of propositions $\alpha, \beta$ for which $P(\alpha/\beta) \neq P(\alpha)$ provided $0 < P(\alpha) < 1$. Obviously such consequences are irrelevant to the evaluation of the non-disjointedness between *A* and *B*. The restriction to relevant consequences forbids this kind of situation by ruling out irrelevant formulas.

The above proposal articulates the notion of 'gerrymandering' or 'ad hocness' in terms of formal relations between the content parts of a hypothesis or a system of hypotheses. If successful, the proposal will hopefully facilitate our search for hypotheses that capture the natural order of things. More relevantly for the purposes of this talk, the proposal is amenable to implementation in the artificially intelligent agents that we hope will be, or at least be among, the scientists of the future.

**References:**

Lakatos, I. (1968) 'Changes in the Problem of Inductive Logic', in I. Lakatos (Ed.), *The Problem of Inductive Logic* (pp. 315-417). Michigan: North Holland Pub. Co.

Schurz, G. (1991) 'Relevant Deduction', *Erkenntnis*, vol. 35: 391 - 437.