Philosophical Engineering: Some Reflections

[Extended Abstract] *

Selmer Bringsjord
Rensselaer AI & Reasoning Laboratory
Department of Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA
http://www.rpi.edu/~brings
selmer@rpi.edu

Categories and Subject Descriptors
Philosophy [Logic]: Robotics

Keywords
philosophy, engineering, deontic logic, robots, ethics

1. INTRODUCTION
Along with a number of colleagues, I have been carrying out AI engineering and software engineering as philosophy. I dub this activity philosophical engineering, and distinguish it from philosophizing about engineering. In the latter mode (one in which we also sometimes engage), one operates as a philosopher, not an engineer; one is doing philosophy of engineering. In the former mode, one is doing engineering and philosophy simultaneously. As an example: The philosopher ponders whether a robot could ever truly be a moral agent, and usually makes a case for a position on the issue. The philosopher-engineer seeks to construct, by leveraging philosophy, a robot that is a moral agent (or perhaps appears to be one, from the standpoint of observable behavior).

2. AN EXAMPLE, BRIEFLY
Many examples of philosophical AI engineering and philosophical software engineering are available in my own case, and in the work of others. To pick one relevant to the present workshop, and in line with what I’ve already said, consider a bit more carefully the attempt to build ethical robots. More specifically, consider the attempt to engineer robots satisfying the following four desiderata.

D1 All actions taken by robots are permissible.
D2 All obligatory actions for robots are performed by them (subject to ties and conflicts among available actions).
D3 No forbidden actions are performed by robots.
D4 All permissible (or obligatory or forbidden) actions can be proved by the robot (and in some cases, associated systems, e.g., oversight systems) to be permissible (or obligatory or forbidden), and all such proofs can be explained in ordinary English.

My approach to engineering robots that satisfy this quartet is a logic-based one, and ethical codes are expressed in logical systems. Details regarding initial efforts in this direction, in which the logical systems in question are proof theories for deontic logic, can be found in (Bringsjord, Arkoudas & Bello 2006). Unfortunately, deontic logics, by definition, only have operators corresponding to the standard ethical categories (obligatory, permissible, forbidden), so a robot built to abide by D1–D4 on the basis of the proof-theoretic side of deontic logic alone couldn’t possess concepts generally agreed to be central to being a moral agent. For example, a robot that knows that it is not lying when uttering some declarative sentence to a human interlocutor must know that it does not intend that causes the human to believe some false proposition. At the workshop, I will describe philosophical engineering devoted to producing a robot that not only satisfies D1–D4, but also has the kind of epistemic attitudes and structures needed to knowingly avoid mendacity. Brief demonstrations of such a robot will be included, through videos.

1In the case of others, a nice example is the philosophical AI engineering of John Pollock (Pollock 2001, Pollock 1992, Pollock 1995). An example in my own case (other than the intersection of robotics and ethics touched upon momentarily) was the attempt to engineer a software system capable of automatically writing stories on the strength of philosophical analysis of literary themes such as betrayal and self-deception. See (Bringsjord & Ferrucci 2000).

2In some formalizations, D3 would be entailed by D1.

3This engineering will exploit innovative efforts by Micah Clark, who at the last North American Computing and Philosophy Conference presented an implementation of definitions of lying proposed by a number of philosophers. Clark’s dissertation, now in progress, is devoted to building a “lying machine.”
3. WHAT EXACTLY IS PHILOSOPHICAL ENGINEERING?

I must confess that while I routinely engage in philosophical engineering, I don’t yet have a philosophically rigorous account of what it is I am doing. What is it to do philosophical engineering, rather than philosophy of engineering, exactly? There is also the general form of the question: What is it to do philosophical X, rather than philosophy of X? Once one tries in earnest to carefully answer these questions, one is engaging in the philosophy of philosophical engineering (and in the general case, the philosophy of philosophical X). One important feature that seems to be necessary for philosophical X is that the key formalisms and frameworks in philosophical X must be central to philosophy itself. In the aforementioned example (engineering ethical robots), the key formalisms have been produced by philosophers, and, at least for the most part, it continues to be philosophers and philosophers alone who refine and extend them. As to a second necessary condition, it seems to me that for philosophical X one or more of the core techniques in philosophy must be central to the process. In the case of my attempts to engineer story generation systems (see note [1]) able to automatically produce narrative involving betrayal, the technique of philosophical analysis (propose a definition, search for counter-examples, adjust the definition if one is found, and iterate; a master of this technique was Chisholm, e.g., see (Chisholm 1976)) was crucial to the attempt to devise a rigorous definition of betrayal.

4. REFERENCES


URL: citeseer.ist.psu.edu/pollock92how.html