SOTO-MORETTINI
Name
Uni

Reverse Engineering the Human: Artificial Intelligence and Acting Theory

> ***"Theatre actors have been staging artificial intelligence for centuries. If one shares the view that intelligence manifests in behaviour, one must wonder what lessons the AI community can draw from a practice that is historically concerned with the infusion of artificial behaviour into such vessels as body and text….Therefore, acting methodology may hold valuable directives for designers of artificially intelligent systems."*** [i]

In an unpublished paper, widely available on the internet, Artificial Intelligence/Robotics researcher Guy Hoffman takes as a starting point that actors have been in the business of reverse engineering human behaviour for centuries. In other words, actors work from observable behaviour (expressed through a written text) backwards, to discover motivation, intention, desire, etc. Of course an actor cannot consider the imagined intentional states without an accompanying consideration of the other factors, such as environment, and the actor's work takes account of the social/human forces that will affect decisions, and determine social 'display rules' in terms of just how much of the character's 'inner state' can or will be displayed. Still, for all this complexity, the actor is in the business of analysing human intelligence and in manifesting intentional states through behaviour, and this makes the area of acting theory (AT) an interesting one in relation to theories and practice in artificial intelligence (AI).

Specifically, Hoffman's paper addresses three areas: (1) psycho-physical unity, (2) mutual responsiveness, and (3) continuous inner monologue. In a later paper specifically focused on Human Robotic Interaction (2011)[ii], Hoffman narrows this down to two areas: continuity and responsiveness. In both papers he references acting theorists (Stanislavski, Sonia Moore, Michael Chekhov, Augusto Boal, Sanford Meisner) and makes specific recommendations based on AT.

In this paper, I want to look at 3 primary questions, in the hope of framing a response to Hoffman's papers:

1) How are the problems of training a human to simulate a human both similar and different from training a machine to simulate a human?
2) How are the problems of AI design similar to the problems that still remain within the area AT?
3) As the field advances, what (if anything) can AT learn from the designers of AI?
In order to consider these questions, I'd like to look closely at two areas that Hoffman addresses: 1. embodied cognition (psycho-physical unity) and 2. the location of responsiveness/action choice/ and the problems of 'single agent' design.
Along with looking at these specific areas, I will be tracing the evolution in thought in both AI and AT (specifically, the work of Tadashi Suzuki and Bogart & Landau), arguing that they have followed similar trajectories.

---

[i] Found at http://alumni.media.mit.edu/~guy/publications/HoffmanAI50.pdf
[ii] http://guyhoffman.com/publications/HoffmanRSS11Workshop.pdf