Guy Hoffman - Fall 2003

A Reading of
Daniel C. Dennet's
Three Kinds of Intentional Psychology

By and large, Dennet introduces three ways to think about intentions in conscious agents, with the explicit goal of formulating a way to reduce folk concepts of belief and desire into physiologically based means of description.

Through his rather involved discussion of the three kinds of intentional psychology, Dennet reaches the conclusion that this reduction should be performed in two stages - one from folk psychology to a so-called "intentional system theory", and one from intentional systems to "sub-personal cognitive psychology" (or - as he suggests in one final turn of argument - the other way around, from sub-personal representation to the black-box intentional theory).

While Dennet does not propose any practical guidelines as to how to perform these reductions, or even how to begin doing so, he goes into length in describing the three different approaches. Several interesting points can be noted in this analysis.

Folk psychology of intentions and beliefs, by far Dennet's most extensively covered type of intentional psychology, is interesting to us in search of a theory on how people perceive intentions and beliefs in artificially created collaborators. His claim that "beliefs are information-bearing states of people that [...] lead to intelligent action" summarizes the important connection between mental constructs and behavior as viewed by the behaviorist school of thought, with which i largely agree. And, since folk are behaviorists, we should provide them with this sort of appearance. In other words, in designing artificial collaborators, we need to create behavior that seems to result from the agents' beliefs and desires to imply existence of the latter.

Dennet also speaks of a basic underlying folk assumption that we view each other as intentional systems and that we assume that our peers' actions are a result of intentions, beliefs and their *rational* derivatives. We can presume that this so-called "approximation" of rationality also holds true of human clients of collaborative agents, and one should aim for this to be a close approximation by mimicking rationality as good as possible.

The paper's description of "internal system theory" I find rather irrelevant. It seems to be a transit point between folk-based behaviorism and true reductionism. This theory can be

grossly described as *labeling internal states without describing their mechanism*. This theory may find its application in the field of comparative studies but not so much in our field of interest, namely constructive theory of mind. I think that viewing the internal representation of intents and beliefs as "black boxes" is not very helpful to our goals.

Finally, the "sub-personal cognitive theory" promises to be the most relevant to our application, and here Dennet lays out the main approach artificial intelligence has used in simulating rationality: "You must put together a bag of tricks and hope nature will be kind enough to let your device get by". While the author claims the above about our own brain (which he calls a syntactical device), it is clear how this applies to even more limited syntactical devices - man-made machines. Unfortunately, his paper does not point us to how to approach this invaluable world of tricks. This leaves us almost where we started.

I find myself hard-pressed to say that "Three Kinds..." displays a firm structure or is easy to follow. Moreover, most of the paper's volume is not supportive of the main argument, but rather a seemingly random collection of thoughts on the three levels of intentional theory, with far too many tangents shooting off from the main point to make it a coherent argument. On the most interesting point, i.e. what approach we might want to take in creating said reduction, I found too little content to render Dennet's analysis useful.