Guy Hoffman - Fall 2003

A Reading of

Bertram F. Malle and Joshua Knobe's

The Distinction between Desire and Intention: A Folk-Conceptual Analysis

In the most empirical of the three papers, Malle and Knobe inquire how people distinguish between an agent's desires and her intentions. Not an attempt at defining these concepts or even at defining their distinction, but rather a folk-psychological research, the paper wishes only to understand how *we* distinguish between the two.

The theory which the authors try to support by ways of experiment and analysis is the following: social perceivers believe that desires lead to intentions on their way to become actions. In order to understand at what point of this process the agent is, these perceivers use three criteria to distinguish desires from intentions:

The content of the pro-attitude. Desires can have any content, whereas intentions usually have only action content, i.e. functional content relating to the agent as the subject of these actions. Additionally, intentions are usually related only to action that the agent believes she can perform.

The role in reasoning of the pro-attitude. Generally, desires function as input of the reasoning process leading from wish to action, whereas intentions are most often the output of this process.

The degree of commitment invested in the pro-attitude. Intentions, due to the persistence (as stated also by Bratman), tend to be viewed by the agent - and thus identified by the social perceiver - as final or almost final, i.e. close to the action. Perceivers measure this degree of commitment by evaluating early investment, invitation of sanctions and acceptance of opportunity costs on the part of the agent.

Malle and Knobe's empirical investigation focuses on phrase frequency analysis and, even more so, on word completion. In trying to show that social perceivers use a particular criterion to classify a pro-attitude, they present the test subject with occurrences of that criterion and then ask them to classify the pro-attitude.

There is a slight methodological shortcoming in this approach: it actually proves the reverse claim. Given a certain measure, this analysis shows that social perceivers will

classify a pro-attitude correctly, but not that these measures are central criteria to the perceivers' everyday classifications.

My second point of criticism is that even were the authors to prove that these three criteria are people's primary instruments of distinction, it is still not very well established how this proves the main theory that folk psychology considers desire and intention as two stages of an "all-things-considered" approach to reasoning. While I do think that this theory is correct, I find the deductive link between the empirical data and the main thesis somewhat weak.

Despite these two structural comments, I still find this paper to be the most insightful of the three. It raises quite a few points that are relevant to the design of cooperative machines:

First and foremost, I believe that the "all-things-considered" tree model of reaching an intent from a number of conflicting desires is a good basis for intent agent design. While this is argued by all three authors, I feel that the description of Malle and Knobe is the most straightforward and practical one. Intent should be the last step before an action, after all other consideration have been weighed as inputs to the reasoning process.

With this in mind, I find it useful to consider "playing out" desires even if they are discarded and don't lead to a particular action. Since desires are perceived by people as preliminary, or *deeper* mental constructs, actually *showing* them in artificial agents will give the illusion of depth and deliberation, and will strongly enforce the identification a client can feel with such an agent.

On another note, the notion of *commitment* is crucial in the design of collaborative agents. If we can build a machine that correctly identifies intents as opposed to desires, this identification should signal a high level of urgency as to the action that is intended. Being able to read the level of commitment in the human agent is a very useful measurement of how immediate the artificial agent should act upon this prospect.

This point is reinforced in the papers insightful final third, in which the authors discuss the function of the analyzed distinction, noting on one hand that if an agent can successfully identify intentions, she can "attempt to predict, explain and *influence* others' actions" and also saying that "intention are more open to debate", because they are almost final.