

A Reading of
Hector J. Levesque et al.'s
On Acting Together

Levesque et al. investigate a formal theory of joint commitment, and later - joint intention and joint action. This formalism is based on the language they introduced for singular goals, intentions and actions.

The first extension over the individual goals and intentions is the introduction of *mutual beliefs*. The main problem with this naïve generalization is that mutual beliefs are by definition symmetrical to the agents. As a result, simply substituting these for individual beliefs doesn't cover the important case in which agent x reaches a conclusion that results in a change of his belief about the status of the goal (i.e. the goal is attained or unattainable), but this conclusion is not shared with agent y.

This is an important point, and Levesque et al. try to resolve it by introducing the notion of a *weak goal of x with respect to agent y* ($WG_x y p$) - which is similar to a regular goal, but once agent x believes that the goal has been achieved, or will never be achieved, he now has a new goal: to create a mutual belief with agent y about this new state of affairs.

This leads us directly to a crucial aspect of collaboration - communication. In order for agent x to fulfill his new goal of changing the mutual belief, appropriate channels of communication must be in place ("if you only told me about it, we could have done something to make it better"). These channels are unfortunately not modeled in this paper.

I see a direct relationship to the argument put forth by Searle in "Collective Intentions and Actions" - namely that social interaction is a pre-requisite for collaboration rather than a result of it.

Another strong point of "On Acting Together" is the formalization of the requirements for collaboration, which might serve as a foundation for an implementation of these principles. Admittedly, it's still a long way from this formalization to implementation.

It should be particularly noted that the described joint rational behavior, based on trusting the other party to communicate a change in plans or sub-plans, in addition to the commitment to communicate your own change of status results in a more robust overall behavior. This should be very interesting to put to empirical testing, examining the practical value of both trust and communication in a multi-agent task scenario.