Lecture 01.07 Robot autonomy and human-robot collaboration

Autonomy is our final essential condition for robots. As with some of the others, it is challenging to draw a line between devices that are and aren't autonomous. Even if we were "hardliners," there would be ambiguity: does autonomy include independence of human influence in *all* things? Consider the following aspects of robot behavior, considered as the behavior of a group or an individual robot:

who the choice of robot(s) acting,
what the action,
when the timing,
where the location,
why the goal, and
how the method.

Any of these six aspects could be autonomous, but it seems too strong to require all these to be autonomous; after all, isn't one of our motivations for making robots having some influence on them? And there are more considerations, such as the programming, construction, and even design of the robot. Perhaps the only fully autonomous robot is an ideal robot approached in an evolutionary process of "alive" (*à la* artificial life) robots.[4] Our perspective is that if any of the above is autonomous, that is sufficient to satisfy the condition of robot autonomy.

It turns out we would like to inhabit the same spaces as robots. In fact, one of our primary motivations for building robots is to have them *help us*. If you've ever been helped by someone over whom you have no influence, you'll start to see the trouble with "fully autonomous" robots. What has proven more valuable in virtually every field of robotics is work that contributes to the better integration of human and robot activities. A way to consider the breadth of this field is to give it two categories.

Human-robot interaction (HRI) is the broad and interdisciplinary study of the interaction of robots with humans, including communication, socialization, and design.

Human-robot collaboration (HRC) is the study of human-robot teams working together to achieve goals. It is sometimes considered a subcategory of HRI.

[4] And even here, one may object that the evolutionary process was started by humans, the conclusion being that a truly autonomous robot is, in fact, impossible.

We will here explore HRC in more detail.

01.07.1 Human-robot collaboration (HRC)

It has been observed that what is hard for a human is easy for a robot and what is easy for a human is hard for a robot. This is often understood as a challenge to human-robot interaction: humans tend to expect robots to be able to perform tasks simple to humans easily, so robots frequently seem downright inept. However, we can also turn this observation around: humans and robots are actually *complementary*. **complementary**

The problem, then, is to find ways for robots to work *collaboratively* with humans—not necessarily replace them. There are, of course, several challenges, most of which have been revealed by attempts to design, build, and deploy collaborative robots. Other challenges were predicted by studying *human collaboration* to discern some essential qualities of collaboration that will likely remain true in successful human-robot collaboration. Before we consider some important ideas that have emerged from these studies, it is worth noting that, since humans find collaboration with humans on many tasks, like moving furniture, simple, we should expect that robots won't. And this turns out to be very much true. **collaboration** **human collaboration**

shared goal Collaboration requires team members share a common goal.

commitment Collaboration requires members of the team to be committed to the achievement of the common goal.

knowledge Collaboration requires members of a team to internally represent knowledge of the states of the environment.

sensitivity Collaboration requires team members be sensitive to the environment, each other, and themselves.

communication Collaboration requires team members to be able to communicate effectively, updating each other about the states of the environment and themselves.

planning Collaboration requires team members be able to plan; that is, to reason through which actions are required to achieve a common goal.

Relatively simple models of practical reasoning (e.g. belief-desire-intention or BDI, Müller (1999)) and relatively detailed cognitive architectures (e.g. Soar, Laird (2012)) have been used to better design robots that can better collaborate with humans.

Several robot control architectures have been developed with insights gained from this work. We will review some important examples in Chapter 04.