Presented at the ONR-UCLA Autonomous Intelligent Networks and Systems Symposium, University of California at Los Angeles, 8-9 May 2002.

Autonomy and Intelligence – A Question of Definitions

Michael R. Blackburn, Ph.D.
Code 2371
Space and Naval Warfare Systems Center
53560 Hull Street
San Diego, CA 92152-7383

(619) 553-1904, mike@spawar.navy.mil

The Problem of Autonomy

It is tempting to call a mechanism that works by itself *autonomous*. The common dictionary definition of autonomy supports this temptation. According to Webster's Ninth New College Dictionary, autonomy is "the quality or state of being self-governing".

But governance implies more than simple perpetual motion. Does the earth's rotation around the sun constitute an autonomous trajectory? Probably we would not commonly think so. What if an asteroid suddenly appeared in the earth's path? What would the earth do about it? Would the earth modify its trajectory and avoid the impending collision? Not likely. A collision would confirm for us that the earth was not autonomous.

Now take a laboratory robot equipped with a SONAR array. The output of the SONAR array could be used to steer the robot away from looming obstacles, thus avoiding many potential collisions. Would we then say that the robot was autonomous? Many robotics developers, based on the titles and texts of their documents, do indeed say that obstacle avoidance is an autonomous behavior. Those very same robots, however, usually have an on-off switch. When that switch is in the off position, the robot cannot avoid a looming object and a serious and destructive collision is likely to occur. The *on* state of the robot could also be compromised by depletion of the robot's energy reserve, again making it vulnerable to environmental calamities. Something similar could happen to us when we are asleep or are otherwise not paying particular attention. Should we say then that autonomy is a state-dependent attribute?

If autonomy is a state dependent attribute, then what mechanisms must an autonomous agent have in place to support its autonomy? From the above example that brought us to this point, we can extract several plausible necessary mechanisms. In no particular order, a short list follows.

- An on-off switch that is always *on*, more or less, but never *off*.
- Sensors to monitor its energy reserves.
- Mobility to acquire new energy and maintain its reserves.
- Sensors for advanced detection of environmental hazards.
- Mobility to protect itself from those hazards.
- Integration (decision) elements that connect inputs to outputs.

It should not be surprising that all animal life have these mechanisms. Therefore, all animals are autonomous. Animals differ, of course, on measures of intelligence, but while living, not on this measure of autonomy. When we consider the necessity of providing for autonomy in artificial agents and we want to be consistent in our definitions, we must prepare for the inclusion of each of the above mechanisms.

Including all of those mechanisms in an artificial agent will not guarantee that our agent will be intelligent, nor long survive in useful tasks under adverse conditions. We should expect this because neither do those mechanisms guarantee the survival of biological agents under adverse conditions relative to the agent's natural ecosystem.

Another interesting fact of the above list of mechanisms is that they are quite interrelated. As energy reserve is key to mobility, and mobility is key both to the maintenance of the energy reserve and to the avoidance of environmental hazards, the autonomous agent must make trade-off decisions for its mobility. Such decisions involve questions about when to move, where to move, and how much to move. The decision-making apparatus is not ordinarily apparent to an outside observer who can appreciate only the external conditions and the resulting behavior of the agent, but the agent is itself aware of both external and internal conditions. The internal conditions that motivate its behavior are primarily related to the state of its energy reserve. Let us call this hunger. Hunger then, in proportion to a scaled response of some intero-sensor, drives the agent toward some source of food that its extero-sensors are appropriately tuned to detect. If however, other extero-sensors detect the presence of hazards, the motor system may be deflected from the perceived gradient of those hazards in proportion to their concentration. The internal decision elements have to weigh these influences in the final output.

The management of energy must be a top priority. Without energy the agent becomes inanimate, and is no more useful than a brick. The avoidance of hazards must also be a top priority for even energy is useless to a disintegrated agent. The foundation for all decision rules then must be the acquisition and management of energy and the avoidance of hazards. In a nutshell, biologists call this criteria *survival*.

Implications for Intelligence

The weighing of costs in a decision making process is greatly improved by the application of intelligence. Intelligence, according to the biological model, is the set of processes that facilitate the energy consumption/expenditure decisions. Intelligence is an exercise in economics.

Intelligent biological systems are trained using rewards, and less so using punishments. Food is a most universal and convenient reward. Analogously, energy should be the reward of preference for artificial agents. Sunlight is a good candidate, but there are others. Robotics developers are aware of the utility of rewards in learning, but often appear to overlook the intervening variable of motivation. A common training paradigm is to reward a robot controller when some consequence of the decision-making algorithm approximates the operator's desired end result, whereupon the operator flips a switch to

signal his/her pleasure at the outcome. Can we say this is reward? To do so would assume that the robotic agent was neither indifferent to the situation that preceded its decision nor to its consequences. That is, the agent had to have some motivation. (The word motivation itself implies a moving force from within.) If rewards, analogous to the biological mechanisms of learning, were to be employed in artificial learning, the robotic agent must be first deprived of some necessary commodity, such as energy, deprived to the degree that its internal sensors for that state quantity were activated and, as a consequence, compelled the robot to move about in its environment in an attempt to restore the depleted quantity. This is all rather elementary, but it is also rather fundamental. If intelligence is a requirement for improved performance in artificial agents, then those agents must first be autonomous, and thus capable of self-motivation.

Implications for Communications

The value of basing a communications network upon autonomous agents of the type defined above may lie in their ability to initiate and persist in certain activities that attempt to improve their own state and that have as well a secondary beneficial effect upon the functionality of the network (our global desired result). Communication difficulties can be solved by the efforts of individual agents to establish links with other agents to meet local needs. While this does not address the issue of the management of information flow across the network, autonomous agents can locally determine the appropriateness of a variety of behaviors to achieve communications, with connectivity as the motivating feedback. These activities could include relocating, mode hopping, and recruiting. The autonomous activities cannot by themselves guarantee that communication will be established and maintained, but if the external conditions permit communications, the viable and motivated agents should eventually achieve that objective. The closer we can associate the functionality of the network with the primary motivations of the agents, the more likely will be the eventual success of our objectives.

Summary and Speculation

The autonomous agent's core objectives are to use its sensors (interoceptors and exteroceptors) and mobility to acquire ambient energy for its own consumption, and maintain its own physical integrity by avoiding threats. It is autonomous only if it can accomplish these objectives. It is, in addition, intelligent only to the degree that it can remain autonomous under adverse conditions. Our concept of an autonomous agent differs from most earlier biomimetic approaches in that the criteria for behavior and adaptation are intrinsic to the agent (not externally imposed). Once the core control foundation is in place, many variations of control architecture may be layered upon it to add capability. Learning is a process that takes advantage of the adaptability of some of these layers. It may also be possible to temporarily suspend the autonomy mechanisms at appropriate times to gain control over such agents and initiate behaviors that benefit ourselves, later to release the agent without permanent architectural damage, but the feasibility of this is speculative. As the fundamental motivation for applying artificial agents to any particular task is to reduce the demands for human involvement, agent autonomy, as we have defined it here, would seem a necessity.