Economically Autonomous Robotic Entities

Carson Reynolds, Alvaro Cassinelli and Masatoshi Ishikawa

***Abstract*— We propose economically autonomous behavior as a novel goal for robotic systems. Currently examples of robotic autonomy are often limited to restricted physical environments such as a factory or road. In this paper we instead restrict the notion of autonomy to a social environment: the economy. We define economically autonomous behavior and describe different levels of independence culminating with hypothetical examples of economically autonomous robotic systems.**

I. INTRODUCTION

It is often the case that a robotics conference one hears presentations about autonomous robots. Such language might amuse or anger the philosophers among us. We might ask: "Autonomous!? Does this mean a robot has free will?" As we shall see autonomy (understood philosophically) and the linked notion of free will are difficult for robots to fully achieve. This paper will define a new variety of autonomy that is perhaps more immediately realizable.

Economic autonomy is the ability to independently operate as a result of income generated. Modern robotic systems need resources in order to function, such as power to drive motors or consumable parts like tires. For economic autonomy it is necessary to obtain resources needed for continuing existence. However, to be independent (in the sense implied by autonomy) these resources may not be the gifts of an interested party (analogous to the food and shelter a parent gives to a dependent child). The resources should instead be acquired with the income generated by the robot's activities.

II. AUTONOMY

Let us compare this definition of economic autonomy to descriptions of autonomy that occur in the literature concerning autonomous robots.

A. Robotic

Much of the work discussing autonomy in robotic systems focuses on avoiding human intervention. Especially in hazardous contexts (such as space exploration or high speed manufacturing plants) it is desirable for a robot to be able to function without the intervention of operators.

In "Animal Behavior as a Paradigm for Developing Robot Autonomy" Anderson and Donath discuss ethological research as well as some different definitions of autonomy in robotics. They describe "self autonomy" as "behavior which may be characterized as supporting self survival." They also discuss "imposed autonomy" as "behavior which does not benefit the robot but fulfills some desired task which we impose upon the system" [2].

It is interesting to draw analogies between self-survival and self-interest in an economic sense. Indeed Adam Smith's theory of economics was predated by an account of the role of self-interest in morality [17].

Franklin and Fraesser provide "A Taxonomy of Autonomous Agents" that applies to biological agents, computational agents and robotic agents [8]. They do so by reviewing a number of definitions for a variety of agents. For instance, they examine Maes' description of autonomous agents as "computational systems that inhabit some complex dynamic environment, sense and act autonomously in this environment, and by doing so realize a set of goals or tasks for which they are designed" [12]. Here Maes' definition mirrors the above description of imposed autonomy. Franklin and Fraesser move on to provide their own definition:

> An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.

They also more narrowly describe autonomy as the property meaning an agent "exercises control over its own actions."

Currently, robots exhibit control over their own actions but only in limited domains. A robotic system might operate in a factory free of direct human intervention. In this situation it has its power provided and is placed in a sheltered environment with a limited task to accomplish to ensure that it can go about existing. A robotic vehicle might take part in an autonomous robotics challenge like DARPA's yearly Grand Challenge. For the duration of the contest the robot would be (figuratively speaking) on its own. The rules of the DARPA's grand challenge in fact stipulate that the "vehicles must demonstrate fully autonomous, unmanned, and safe operation" [1].

However, the ideal of full autonomy (that is autonomy as exhibited in all spheres of human life) is still not a capability that robots exhibit. We would argue that this is due in part to the complexity of tackling autonomy in the sense implied by philosophy.

B. Philosophical

Autonomy is an often invoked but controversial element in philosophy. It plays a fundamental role in both political philosophy (specifically liberalism) as well as moral philosophy [5].

Kant expounds on autonomy directly and at length:

This research is partly supported by Special Coordination Funds for Promoting Science and Technology, "IRT Foundation to Support Man and Aging Society."

All authors are with the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan carson@k2.t.u-tokyo.ac.jp

> Autonomy of the will is the property of the will through which it is a law to itself (independently of all properties of the objects of violation). The principle of autonomy is thus: 'Not to choose otherwise than so that the maxims of one's choice are at the same time comprehended with it in the same violation as universal law.' That this practical rule is an imperative, i.e., the will of every rational being is necessarily bound to it as a condition cannot be proved through the mere analysis of the concepts occurring in it, because it is a synthetic proposition; one would have to advance beyond the cognition of objects and to a critique of the subject, i.e., of pure practical reason, since this synthetic proposition, which commands apodictically, must be able to be cognized fully *a priori*; but this enterprise does not belong in the present section. Yet that the specified principle of autonomy is the sole principle of morals may well be established through the mere analysis of the concepts of morality. For thereby it is found that its principle must be a categorical imperative, but this commands neither more nor less than just this autonomy [10].

While Kant's description is characteristically circuitous we can see that for Kant, autonomy acts as "the sole principle of morals." Self-imposition of autonomy in pursuit of the universal moral law was for Kant a crucial process for moral beings.

While Mill does not explicitly use the word autonomy, one can see the value supported by his vigorous defense of liberty [14]. Mill's championing of individual liberty and bounds for political authority (we argue) are evidence of tacit approval of autonomous behavior. Moreover, Mill's variety of utilitarianism implicitly invokes the individual's choice and argues that it should be for higher pleasures as opposed to physical pleasures.

Indeed for moral philosophers autonomy becomes a value of central importance because if we cannot choose our actions then discussions about accountability and blame become less coherent. However (especially in the case of robotics) it may not be possible to assume the freedom of will required for full autonomy.

III. Free Will & Robotics

Is it possible for a robot be autonomous in the philosophical sense? Can a robot have its own intent? When a robot chooses, does it not do so in a manner prescribed by its construction?

These questions are usually debated in philosophy as problems of concerning free will. If a robot can control its decisions then it has free will. And likewise, if a robot can control its decisions then it is also possible for the robot to exercise "control over its own actions" and thus be autonomous according to Franklin and Fraesser's definition.

What obstacles could exist to a robot being fully autonomous or having free will? From a naive point of view there are two classes of obstacles: obstacles that apply to all entities and those that are robot specific.

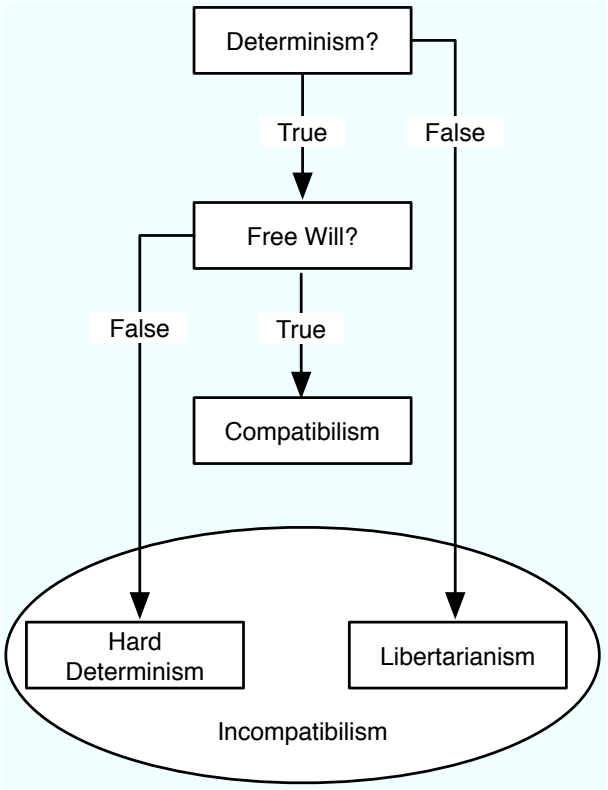

Fig. 1. Hubbard's taxonomy of positions concerning free will [9].

In the case of universal obstacles to free will there are a number of positions one might take. Hubbard provides a convenient taxonomy of major positions that is reproduced in figure 1.

Some philosophers deny that free will exists at all. Often some variety of determinism is used to justify this position. These philosophers are labeled "hard determinists" for their complete embrace of determinism and complete rejection of free will. For instance, those who subscribe to physical determinism may argue that a robot's actions are already determined by the dynamic properties of the physical universe. We see an extreme version of this position echoed by the physicist Laplace:

> We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it–an intelligence sufficiently vast to submit these data to analysis–it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. [6]

This intelligence, which is colloquially referred to as

Laplace's demon, is an allegorical representation of physical determinism. So those who strongly commit to Laplace's view think of the universe as a system that (if all relevant information is known) will precisely describe past and future. In such a world, all past present and future robots have no actual autonomy. Their behaviour is already dictated by the state of the universe.

A problem more closely related to robotics is that of deterministic computation. Ironically (for our current subject matter) a great deal of effort is expended in designing various computational elements to be deterministic. If we use a Turing machine as a model for computation then we can define deterministic to mean "permitting at most one next move at any step in a computation" [3]. This effectively means that there is no unpredictable behavior among deterministic computational systems. Some argue that a system (such as a robot) that makes use of deterministic computation cannot exhibit free will [15].

It should also be noted that there are non-deterministic computational schemes. Some such schemes are simply probabilistic (which may make those seeking intentionality in the will-choosing process to reject them) [4]. Other non-deterministic systems are composed of parallel processes, each of whose behavior is itself deterministic.

Others whose position is labeled compatibilist (meaning that free will and determinism are compatible), believe that the physical world is deterministic but there is still some room for free will. See Dennett's Elbow Room for a more articulate and complete version of this argument [7].

Following Dennett's line of thought it seems that much of the work for roboticists seeking free-will-like behavior is in solving the problems in creating an object which projects enough evidence for agents to assume an intentional stance towards it. Put more simply, a robot needs to exhibit some coherent reasonable behavior in choosing. If it just chooses probabilistically we have no more reason to believe the robot has free will than we do a slot machine or pair of dice. However, if the robot has no intent of its own (by just doing as its designer's plan) different problems arise. So, we may speculate, the problem lies in how the robot chooses (or evolves) its intentional behavior.

McCarthy adopts the compatibilist position and applies it to robotics. McCarthy argues that free will can be achieved by "internal decision processes even if these processes themselves are deterministic" [13]. He presents a finite state automata system whose behavior is deterministic but arrives at its decision by examining internal states (such as the result of a common-sense reasoning process). For McCarthy the ability to behave freely is embodied in the phrase "I can but I won't" and he further argues that his finite state automata mimic the linguistic behavior of "can." He asserts that this is a variety of free will.

IV. The spectrum of economic autonomy

Now that we have reviewed both robotic and philosophical definitions of autonomy it is time to return to the main topic of the paper: economic autonomy.

Imagine that a robot is able to earn money by performing work. This is not unreasonable. For instance, industrial robotics companies charge companies right now tens of thousands of dollars for robotic arms. However, imagine that the robot's earnings did not go to the manufacturer but instead were the robot's own.

How might this come about? A robot would certainly owe some resources to its manufacturer for the cost of designing and manufacturing it. However, why should robots be perpetual slaves as opposed to indentured servants for a set time?

Let us suppose that a robotic designer is interested in his robotic systems having full autonomy. Why would such an expectation stop with respect to the robot's finances? Obviously full autonomy encompasses the financial domain. Why would the sympathetic designer not allow the robots they create to become financially autonomous?

So let us describe a spectrum of different levels of autonomy bounded at one end by parasitic robots and the other end by a robots capable of reproducing themselves through their own financial wherewithal.

In the extreme case a robot is not economically autonomous so much though that it acts as parasite feeding on the money and resources of roboticists. Some researchers have taken this idea so far as to develop a parasitic humanoid which relies on a human wearer for both mobility and training information [11]. Conversely, in the media-art installation maschine-mensch, the humans are controlled by the machine they themselves created, and reduced to slavery as part of a larger mechanical system [18].

A more moderate case would be a robot that generates income which is applied to its upkeep but only a portion of its costs. So for instance if a robot generates income through entertainment performance but this income does not cover the costs associated with manufacture and maintenance then it is this more moderate case.

One can also imagine a robots that metaphorically "pays for itself." This robot earns enough income to pay for its design manufacturer and upkeep. Existing industrial robots would probably fit into this class if they were able to keep their own earnings. Presumably the robots perform enough work to justify their purchase. If such a robot were legally entitled to keep the proceeds from such work it could conceivably emancipate itself (financially at least). However, currently in countries in which the legal code give a privileged position to natural persons, the robot seems to be relegated to the role of property (which could seized and made to work for purposes contrary to the robot's interests). Indeed there are many interesting questions related to property and autonomy that economically independent robots could force upon the legal system. But let us continue along the our original plan and further describe the spectrum of economic autonomy.

Further along we can imagine a robot which earns enough income to both pay for itself and pay for parts with which to repair itself. This possibility is intriguing in that the robot could "live" indefinitely (supposing that someone exists who

is willing to make parts for what the robot can pay). Such a robot has a variety of subsistence economic autonomy where it is able to survive and keep at the level of functionality roughly equivalent to that at the time of its manufacture. A trivially more sophisticated economically independent robot would be one that is able to earn enough income to improve its capabilities.

Toward the far end of the continuum we can hypothetically conceive robots that can earn enough money to fund their reproduction. Taking an mutualist view (such as Pollan did in Botany of Desire (where he argued that plants exploit us to grow and improve themselves)) one might argue that automated computer trading systems whose success begets bigger and better automated trading systems exhibit this sort of behavior [16]. So let us imagine that a peculiar sort of auto-mat opens up on the street corner. This auto-mat operates like most others dispensing goods with a robot arm. However this robotic auto-mat uses the profits from its sales to pay for the manufacture and assembly of copies of itself. As such it provides an example of economic autonomy.

An amusing permutation of this line of thought is one possible shortcut to legal person-hood for autonomous robotics. If a future robot were able to show that it is functionally equivalent to a natural person (in the jurisprudence sense) then conceivably a group of robots could found a corporation that would have some civic rights normally reserved to persons (such as filing lawsuits).

V. Experimenting with Perceived Autonomy

Do we view artifacts that exhibit autonomy as being more accountable and appropriate to blame for wrongdoings? This forms an interesting and experimentally testable hypothesis about how autonomy (economic or otherwise) might interact with moral notions like blame, harm, and accountability. An ideal design for an experiment would be to present participants with a variety of robots exhibiting different sorts of autonomous behavior and then to ask participants to ascribe intentions to the robots.

This paper is a collection of preliminary ideas for both creation of economically autonomous robots and examining how they are viewed by other moral beings.

References

[1] Defense Advanced Research Projects Agency. Urban challenge: Rules, December 2006. http://www.darpa.mil/grandchallenge/docs/Urban_Challenge_Rules_121106.pdf.

[2] T. L. Anderson and M. Donath. Animal behavior as a paradigm for developing robot autonomy. In P. Maes, editor, *Designing Autonomous Agents*, pages 145–168. MIT Press, Cambridge, MA, USA, 1994.

[3] Mikhail J. Atallah, editor. *Algorithms and Theory of Computation Handbook*. CRC Press LLC, 2007. http://www.nist.gov/dads/HTML/deterministc.html.

[4] Andrea Bianco and Luca de Alfaro. Model checking of probabalistic and nondeterministic systems. In *Proceedings of the 15th Conference on Foundations of Software Technology and Theoretical Computer Science*, pages 499–513, London, UK, 1995. Springer-Verlag.

[5] J. Christman. Autonomy in moral and political philosophy, 2003. http://plato.stanford.edu/entries/autonomy-moral/.

[6] Le Marquis Pierre Simon de Laplace. *A Philosophical Essay on Probabilities*. Dover Publications, January 1996.

[7] Daniel C. Dennett. *Elbow Room: The Varieties of Free Will Worth Wanting*. The MIT Press, November 1984.

[8] Stan Franklin and Art Graesser. Is it an agent, or just a program?: A taxonomy for autonomous agents. In *ECAI '96: Proceedings of the Workshop on Intelligent Agents III, Agent Theories, Architectures, and Languages*, pages 21–35, London, UK, 1997. Springer-Verlag.

[9] E. M. Hubbard. Free will taxonomy (graphic) — wikipedia, the free encyclopedia, 2007. http://en.wikipedia.org/w/index.php?title=Free_will&oldid=101180634.

[10] Immanuel Kant. *Groundwork for the Metaphysics of Morals*. Yale University Press, November 2002.

[11] T. Maeda, H. Ando, M. Sugimoto, J. Watanabe, and T. Miki. Wearable robotics as a behavioral interface - the study of the parasitic humanoid. In *Proceedings of Sixth International Symposium on Wearable Computers, 2002. (ISWC 2002)*, pages 145–151, 2002.

[12] Pattie Maes. Artificial life meets entertainment: lifelike autonomous agents. *Commun. ACM*, 38(11):108–114, November 1995.

[13] J. McCarthy. Free will - even for robots. *Journal of Experimental & Theoretical Artificial Intelligence*, pages 341–352, July 2000.

[14] John S. Mill. *On Liberty*. Penguin Classics, July 1982.

[15] Roger Penrose. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics (Popular Science)*. Oxford University Press, March 1999.

[16] Michael Pollan. *The Botany of Desire: A Plant's-Eye View of the World*. Random House Trade Paperbacks, May 2002.

[17] Adam Smith. *The Theory of Moral Sentiments (Great Books in Philosophy)*. Prometheus Books, May 2000.

[18] T Zucali and C. Rhomberg. maschine-mensch, 2006. http://maschine-mensch.net/.