Penultimate self-produced version.

Published in the journal *Paradigmi* **33**(3), pp. 163-178.

DOI: 10.3280/PARA2015-003010

GUGLIELMO TAMBURRINI*

MACHINE INTELLIGENCE SPORTS AS RESEARCH PROGRAMS

Abstract: Games and competitions have played a significant role throughout the history of artificial intelligence and robotics. Machine intelligence games are examined here from a distinctive methodological perspective, focusing on their use as generators of multidisciplinary research programs. In particular, Robocup is analyzed as an exemplary case of contemporary research program developing from machine intelligence games. These research programs arising are schematized in terms of framework building, subgoaling, and outcome appraisal processes. The latter process is found to involve a rather intricate system of rewards and penalties, which take into account the double allegiance of participating scientists, trading and sharing interchanges taking place in a multidisciplinary research environment, in addition to expected industrial payoffs and a variety of other fringe research benefits in the way of research outreach and results dissemination, recruitment of junior researchers and students enrollment.

Key words: double allegiance of scientists, machine intelligence games, methodology of AI, methodology of robotics, multidisciplinary research, research gamification.

Parole chiave: doppia fedeltà degli scienziati, macchine intelligenti e giochi, metodologia dell'IA, metodologia della robotica, ricerca multidisciplinare, ludicizzazione della ricerca.

1. Introduction

Games and competitions have played a significant role throughout the history of robotics and artificial intelligence (AI from now on). Pioneers of AI – including Alan Turing, Claude Shannon, John McCarthy, Allen Newell and Herbert Simon – identified chess as a fruitful domain for AI inquiry and contributed to early computer chess research (Cordeschi, 2002, pp. 177-179). Nowadays, Robocup and the DARPA Robotic Challenge are ongoing robotic competitions involving research groups worldwide; the entry "Games and puzzles" figures in the top level list of AI research topics compiled by the AAAI (Association for the Advancement of Artificial Intelligence; http://www.aaai.org/home.html); and international funding agencies make ample provisions for robotic competitions.

A chief motive of methodological interest for machine intelligence games arises from clear rules of behavior and success criteria that enable one to address, and in some cases to sidestep successfully, conceptual and empirical difficulties besetting the definition and the investigation of other sorts of intelligent behavior. Moreover, game task environments are suitable to investigate central AI and robotics issues, such as action planning and coordination, heuristic search, and knowledge representation. Competitions based on social games have additionally provided significant test beds for measuring and ranking the playing performances of AI and robotics systems resulting from the integration of various skills and capabilities.

Machine intelligence games and competitions are examined here from yet another methodological perspective. Their role as *generators of research programs* is discussed and analyzed on the basis of the processes of *framework building*, *subgoaling*, and *outcome appraisal*. The first two of these processes –

* DIETI – Department of Electrical Engineering and Information Technologies, Università di Napoli Federico II, Naples, Italy; e-mail: tamburrini@unina.it. This article takes its origin in extensive discussions with Roberto Cordeschi during the winter 2013-14, which concerned both structure and distinguishing features of research programs in AI. I am grateful to Thomas Christaller for stimulating methodological discussions on RoboCup. All webpages mentioned in the text were last visited on October 5th, 2015.

framework building and subgoaling – are found to introduce relatively modest variations on similar processes described in traditional two-process models of scientific inquiry (Godfrey-Smith, 2003), which notably include Thomas Kuhn's model of normal science (Kuhn, 1962), Imre Lakatos's model of scientific research programmes (Lakatos, 1978), and Larry Laudan's model of research traditions (Laudan, 1977). However, the outcome appraisal process of a machine intelligence research program based on games – which involves periodic evaluations of the various results that the research program brings about – affords a more interesting case study for the methodology of scientific and technological inquiry. This circumstance arises especially in view of the fact that research programs based on games are markedly multidisciplinary in character, and require the coordinated participation of AI and robotics research communities which are dispersed over a wide range of competences, research goals and expectations, methodologies, and even their preferences in the way of software platforms, robotic platforms and other technical tools. Correspondingly, we analyze here the rather intricate system of regional rewards and penalties, multidisciplinary trading and sharing interactions, and other fringe benefits that must be taken into account to appraise progress and to appreciate the contributions, if any, that these research programs are making to scientific and technological advancements. This system of rewards is exemplified by means of the robotic soccer playing competition Robocup (Asada *et al*., 1999; Kitano *et al*., 1997), which is used as an exemplary case of research program developing from machine intelligence games.

Roberto Cordeschi's comprehensive analysis of the beliefs, aspirations, and research *themata* that coexist in the cluster of AI and robotics research communities (Cordeschi, 2002, pp. 241-279) is a key to understanding the wide range of motivations and expected rewards leading heterogeneous groups of scientists to converge on machine intelligence games research. Usually, AI and robotics scientists expect to gain, from their secondary allegiance to machine intelligence inquiries on games, insights and results that are significant from the standpoint of their primary allegiance to well-established research areas in AI and robotics. Thus, the outcome appraisal process for machine game inquiries requires an understanding of the *double allegiance* of participating scientists and the regional system of rewards associated to their primary allegiance.

Of particular interest for the analysis of secondary allegiances to machine game inquiries is Cordeschi's analysis of the motives underlying the initial involvement and the ensuing withdrawal from computer chess investigations of scientists focusing on human problem solving and its computer simulation. He pointed out that the use of more powerful hardware resources and statistical learning methods, while improving computer chess performances, progressively weakened the secondary allegiance of scientists expecting to gain from chess programs significant insights into the heuristic problem solving strategies of humans and other cognitive systems (Cordeschi, 2002, p. 178 and p. 182; Cordeschi and Tamburrini, 2015).

Secondary allegiances to research programs on machine intelligence games are additionally motivated by expectations of fruitful trading and sharing activities occurring in multidisciplinary research environments. Furthermore, the integration of machine intelligence technologies and systems which is needed to cope with game playing and competitions is expected to feed the pipeline between academic research and industry. And finally, the gamification of machine intelligence research is usually expected to bring with it various fringe benefits, by motivating junior researchers, attracting wider cohorts of science and engineering students, and raising the awareness and interest of the general public towards AI and robotics.

On the whole, machine intelligence games afford an interesting vantage point from which to explore the rich network of regional interests, rewards and diminishing returns, trading and sharing interactions, connections between academic research and industrial R&D, in addition to the ancillary motivational and communication benefits that jointly contribute to determine the waxing and waning of multidisciplinary research programs in machine intelligence and in other fields of research. To begin with this exploration, let us turn to consider RoboCup's long-term objective of robotic soccer teams competing on equal terms with the best human soccer teams.

2. RoboCup's regulative idea

When RoboCup started in 1997, the conjecture was initially advanced that a team of fully autonomous robots would eventually defeat the human world champion soccer team by the year 2050. Later versions of the RoboCup manifesto more cautiously suggested that the goal of beating the best human soccer team «will take decades of efforts, if not centuries. It is not feasible, with the current technologies, to accomplish this goal in any near term.» (http://www.robocup.org/about-robocup/objective/). If robotic soccer teams are expected to perform poorly against human teams for many years to come, then RoboCup's ultimate goal is neither a realistic objective for the current generation of robotic engineers to pursue nor a useful benchmark to assess progress in robotic soccer playing from one RoboCup tournament to the next. Therefore, one may legitimately wonder why this elusive target was proposed at all, and whether it has any effective role to play in the context of current robotics research.

The RoboCup manifesto addresses these concerns suggesting that the visionary goal of beating the human world champions enables one to shape a research agenda formed by "a series of well-directed subgoals." A reflection on early RoboCup activities suffices to illustrate this generative role. When RoboCup started, robotic technologies were sufficiently advanced to build robotic teams playing rudimentary soccer games. Measured with the yardstick of FIFA world championship success, the performances of these robotic players were totally unsatisfying. Nevertheless, their playing levels made initial benchmarks available to evaluate the next generation of teams participating in RoboCup leagues. Similarly, the improved skills of later generations of robotic teams made new benchmarks available for further research in the various league. These playing benchmarks were supplemented with more traditional research targets identified in RoboCup scientific symposia and concerning specific skills making soccer playing possible, such as task-oriented visual feature selection, behavioral pattern modeling and coordinated action planning.

There is no end in sight for improving robotic soccer capabilities using these various benchmarks from the vantage point of good human soccer playing levels. Hence, the RoboCup long-term goal is sensibly construed as an organizing or regulative idea for inquiry. This regulative idea is productive of a series of meaningful scientific and technological subgoals to pursue, independently of whether and when a robotic soccer team will eventually beat the soccer world champions.

RoboCup pioneers suggested a simile between the productive role of their visionary sports objective and the structure of the space program that President John F. Kennedy announced in a 1961 address to the US Congress by reference to the ultimate goal of «landing a man on the Moon and returning him safely to the Earth.» The accomplishment of relatively more modest subgoals – from the development of suitable fuel and missile technologies to medical interventions supporting human life in microgravity conditions – enabled one to fulfill prerequisites for manned missions to the Moon. Equally important, the achievement of these subgoals made a variety of new scientific knowledge, technologies and systems readily available to industry and other academic research areas. Accordingly, the ambitious long-term goal of the US space program in the 1960s fertilized many areas of science, technology and industrial R&D before and independently of the walk that astronauts Armstrong and Aldrin took on the lunar surface on July 20th, 1969.

Three different processes are clearly identifiable in the above narratives about RoboCup and the Apollo program: *Framework Building*, *Subgoaling*, and *Outcome Appraisal*. Framework Building is the process of identifying long-term goals that are well beyond state-of-art in both science and technology, along with scientific and technological tools that are initially suggested as appropriate means to pursue those goals. Subgoaling is the process of identifying and pursuing a series of increasingly ambitious objectives that are coherent with the general framework. Unlike the long-term goals of the research program, subgoals are expected to be both achievable and rewarding in the short term. Outcome appraisal is the process of identifying and periodically applying suitable evaluation criteria that enable one to appreciate the contributions, if any, that the research program is making to science, technology and industry by pursuing its feasible subgoals.

The two processes of framework building and subgoaling that are involved in machine intelligence research programs on games appear to simple variations on processes that are variously described in two-process models of *scientific* inquiry (Godfrey-Smith 2003, pp. 117-121). Influential models in this class are Thomas Kuhn's model of normal science functioning, Imre Lakatos's model of scientific research programmes, and Larry Laudan's model of research traditions. Indeed, the establishment of a Kuhnian paradigm provides a general framework for the daily puzzle solving activities of scientists; the hard core of a Lakatosian scientific research programme is a framework guiding the search for a series of models, which jointly act as a protective belt around the hard core, provided that each model in the series is characterized by increasing predictive and explanatory powers; and Laudan's evolving research traditions, with their sets of methodological and ontological commitments, provide revisable frameworks orienting scientists in their selection of scientific allegiances and their choice of problems to solve.

The outcome appraisal process of a machine intelligence research program based on games, with its periodic evaluations of results that have been obtained, affords a more interesting case study for the methodology of scientific and technological inquiry. The markedly multidisciplinary character of RoboCup and other research programs on machine intelligence games, their combination of scientific, technological, and industrial objectives, in addition to their seeming lack of useful technological applications suggests that an intricate system of rewards and penalties must be taken into account in any sensible attempt to identify and appraise their outcomes.

3. Double allegiance and regional rewards

The development of robotic soccer players requires converging contributions and provides an integrated task for a variety of machine intelligence research areas, such as «real-time sensor fusion, reactive behavior, strategy acquisition, learning, real-time planning, multiagent systems, context recognition, vision, strategic decision-making, motor control, intelligent robot control, and many more.» (http://www.robocup.org/about-robocup/objective/). Accordingly, identifying what RoboCup is contributing to each and every one of these research areas naturally suggests itself as a sensible criterion to use in order to appraise the outcomes of this research project. Indeed, qualitative estimates of Robocup contributions to regional objectives of AI and robotics research have been actually provided by reference to, e. g., techniques for team coordination and distributed planning, temporally constrained sensor fusion and visual processing, motion detection and modeling, efficient solutions of optimization problems with uncertain and incomplete data (Visser and Burkhard, 2007). In their analysis of these regional contributions, Visser and Burckhard point out that scientists participating in RoboCup usually combine this research activity with other machine intelligence inquiries: «Most of the teams combine their work with other challenges from outside RoboCup in terms of addressing scientific questions as well as technical solutions. The evaluation and the exchange of ideas during the competitions and the symposium is an effective and worthwhile foundation for their future work.» (Visser and Burckhard, 2007, p. 132).

This combination of scientific aims and activities reflects the *double allegiance* of most scientists participating in a multidisciplinary research program like RoboCup. Their professional identities, career expectations and research ambitions are not centered around RoboCup. Rather, these crucial properties of their profile are defined by selective reference to the entrenched AI and robotics research areas that are mentioned at the beginning of this section. Accordingly, starting from a primary allegiance to one or more traditional research areas in machine intelligence, these scientists endorse robotic soccer playing inquiries as a secondary allegiance only. More specifically, they pursue soccer sports objectives chiefly in view of on expected impact on traditional research areas in AI and robotics. And their continued commitment to RoboCup and other research programs based on machine intelligence games typically depends on the fruitfulness of their secondary allegiance activities from the standpoint of their primary allegiance. This means that various appraisal criteria converging on an evaluation of RoboCup subgoaling activities are inherited from the primary allegiance of contributing scientists. And these imported appraisal criteria profoundly influence the individual scientist decision to whether renew or else to waive her secondary allegiance commitments.

The history of computer chess affords a significant illustration of how appraisal criteria reflecting the double allegiance of participating groups of scientists affects their commitment decisions and the evolution of multidisciplinary research programs on machine intelligence games. Specifically, diminishing returns for primary allegiance research areas resulted in selective breakaways from computer chess research at a time when the performances of computer chess programs were dramatically improving.

To illustrate these dynamical changes in computer chess inquiries, let us recall Cordeschi's careful analysis of the motives driving towards computer chess inquiry *theoretically oriented* AI scientists, that is, scientists who were primarily interested in understanding computational principles and mechanisms enabling both biological systems and machines to manifest intelligent behaviors (Cordeschi, 2002, p. 171 and pp. 176-179). Allen Newell and Herbert Simon designed programs to simulate human cognitive processes underlying chess playing behavior. In accordance with the basic tenets and goals of their Information Processing Psychology (Cordeschi, 1984, p. 341 ff.), Newell and Simon prized computer simulations of human chess playing as sources of psychological explanations of the underlying mental processes. John McCarthy championed a non-simulative approach in theoretically oriented AI. Nevertheless, he engaged into computer chess inquiries insofar as chess playing strategies could be applied by cognitively or computationally limited systems, of both biological and artificial *genera*, to cope successfully with a wide variety of complex information processing tasks (see Cordeschi 2002, pp. 200-201). Clearly, in either one of its simulative and non-simulative variants, theoretically oriented AI did not prize tournament victory in and for itself.

AI scientists pursuing computer chess research from an engineering (rather than from a theoretically) oriented perspective prized tournament victory and found, on this account, good motives to relinquish in their work the simulation of human problem-solving strategies. In a panel on computer chess held at the

1991 International Joint Conference on Artificial Intelligence, Feng-hsiung Hsu, one of the designers of the chess program Deep Thought, pointed out that inputs from human chess experts cannot be invariably trusted for the purpose of improving computer chess performances. And he went as far as to claim that various changes to the Deep Thought heuristic procedure suggested by human experts had been found to affect negatively its performances. Therefore, he concluded, human expert inputs have to be limited and relevant knowledge for computer chess playing should more effectively and efficiently generated on a statistical basis rather than relying on symbolic learning or explicit domain modeling (Levinson *et al.*, 1991, p. 548). This design approach played a significant role in Deep Blue's victory over Kasparov in 1997. This result, however, was relatively unimportant, as Cordeschi remarked, from the standpoint of Newell and Simon's simulative approach:

> [T]he defeat of a human being by a computer that plays chess, as in the case of the Deep Blue program that beat Kasparov in 1997, while interesting, is not a particularly significant result from the point of view of those (first of all Simon) who remained interested in the psychological realism of simulation… "Dissimilar processes yield similar conclusions": this is, moreover, the comment of the designers of Deep Blue, who emphasized "the contrast in styles" between man and machine during the match. (Cordeschi, 2002, p.182n.)

John McCarthy was similarly dissatisfied, albeit for different motives, with the growing emphasis on tournament success and the use of more powerful hardware resources in computer chess. He expressed his reservations about these developments in a review of a book devoted to the history of chess programs and their tournament victories. There, McCarthy claimed that tournament success had come with relevant scientific costs. Drawing on Alexander Kronrod well-known simile – according to which "Chess is the *Drosophila* of artificial intelligence" – McCarthy complained that «computer chess has developed much as genetics might have if the geneticists had concentrated their efforts starting in 1910 on breeding racing *Drosophila*. We would have *some* science, but mainly we would have very fast fruit flies.» (McCarthy, 1997, p. 1518). And he went on to argue that computer chess inquiries were no longer a good source of insights into the general principles and mechanisms that enable cognitively bounded agents to cope with computationally complex tasks. In particular, he identified a chief motive for this diminishing interest in the use of increasingly powerful hardware resources and the corresponding possibility of chess programs exploring more and more moves ahead in the game. Chess programs exploiting this feature were better players than earlier chess programs relying on less powerful hardware and more limited search ahead strategies. By the same token, however, the better players use heuristic strategies which fail to shed light on the mechanisms that more limited computational agents use to deal with chess and other computationally complex problems.

In conclusion, reasons to withdraw from computer chess research emerged from both simulative and non-simulative viewpoints in theoretically oriented AI on account of the increasing emphasis on tournament success which impressed a change in algorithmic methods and hardware technologies. From the standpoint of RoboCup there is a lesson to learn from these changing attitudes towards computer chess inquiries. Videos of matches from RoboCup leagues, when presented in their temporal sequence, afford an impressive demonstration of steadily improving robotic soccer capabilities; more quantitative approaches finalized to evaluate game playing progress in the Simulated robotic soccer league point in the same direction (Gabel and Riedmiller, 2012). However, tournament success may not be a significant goal for various participating scientists and may even drive them away if the more promising methods for successful soccer playing are uninteresting or even detrimental from the vantage point of their primary allegiance.

4. Integration and industry, trading and sharing

Rewards that group of researchers may obtain from their participation in RoboCup activities are not limited to those that are especially prized from the vantage point of their primary allegiances. Indeed, the RoboCup manifesto emphasizes the importance of boosting industrial R&D, and endorses the capability of feeding the pipeline between academic research and industry as a major appraisal criterion to adopt. This is a tall order criterion to meet, also in view of the fact that machine intelligence academic research has not invariably produced significant and sustained industrial payoffs. The so-called AI winter of the early 1990s, in the course of which many AI companies closed down after the booming expansion of AI industry in the previous decade (Russell and Norvig, 2010, p. 24), was an acute manifestation of the difficulty to bridge the gap between machine intelligence academic research and its industrial applications.

The heuristic thinking of engineers is a crucial starting point to understand how models, technologies and systems that are initially oriented to robotic soccer playing might be translated into useful real-world applications. Thus, for example, the cooperative behaviors of robotic soccer players are potentially useful for developing teams of robots that are required to coordinate their actions and to perform task assignments in dynamic environments comprising human and non-human agents and involving changing illumination, sound, and surface conditions. These integrated skills are needed by teams of service robots for which markets are steadily expected to grow (http://www.ifr.org/service-robots/statistics/), such as surveillance robots operating on museum grounds and other public areas, logistics and warehousing robots, and platoons of self-driving cars.

The translation problem arising from the absence of a direct connection with real-world applications is not unique to robotic soccer players. One has to address this problem in connection with other striking feats of robotic engineering, such as the charming performances of the humanoid robot ASIMO and cheerfully marching bands of robotic wind instruments players. There is more than meets the eye in these captivating performances. Indeed, the 130 patents that the Honda obtained on account of research work on ASIMO enabled this company to develop useful real-world applications, including assisted mobility devices that are designed to help persons with weakened leg muscles due to age, accidents or illnesses. (http://www.gizmag.com/honda-stride-management-assist/23512/).

Only lists of topics compiled by the AAAI or the technical committees of RAS (Robotics and Automation Society; http://www.ieee-ras.org/technical-committees) have been used so far in this article to classify and distinguish between groups of scientists participating in multidisciplinary machine intelligence research programs based on games. There are, however, many additional properties of research groups one may informatively use for our present methodological concerns. These properties include the robotic platforms employed by each research group, the application domains they are interested in, their preferred research methodologies, and even the supply industries they place their orders to for specific components and subsystems. The research communities that one may identify on these enriched grounds for classification are comparable to a mosaic of subcultures. Interactions within each subculture are typically more extensive than interactions between subcultures. But goals, methods, platforms, and application domains of other subcultures are not completely ignored, insofar as contacts between subcultures pave the way to fruitful transactions.

To identify the mechanisms governing scientific and technological transactions within a markedly varied landscape of research interests, Peter Galison suggested that one may profitably learn «… from the anthropologists who regularly study unlike cultures that do interact, most notably by trade. Two groups can agree on rules of exchange even if they ascribe utterly different significance to the objects being exchanged; they may even disagree on the meaning of the exchange process itself. Nonetheless, the trading partners can hammer out a *local* coordination despite vast *global* differences.» (Galison, 1997, p. 783). Galison himself explored these interactions in the history of physics, by reference to interactions between the partly autonomous practices of instrumentation, experimentation, and theory (Galison 1997).

In RoboCup, exchanges between subcultures are not limited to trading transactions. Also the sharing and standardization of technologies and systems plays a methodologically significant role, insofar as these enable one to overcome or alleviate methodological problems affecting solitary robot construction practices. The behavior of robots constructed in one research lab depends on a wide variety of individual design choices and implementation details which concern, e. g., robot size, sensor systems, selection of actuators and balanced composition of their hardware and software components. As a consequence, one can hardly compare in depth the performances of the resulting systems to the superficially similar performances of robots constructed elsewhere on the basis of different design and implementation choices. RoboCup sharing policies provide some means to go beyond solitary robot construction practices, thereby facilitating more principled comparisons between the performances of robotic systems and their underlying mechanisms. Notably, the practice of making programs publicly available has been encouraged in the Simulation league; and standard industrial platforms have been increasingly adopted in both Simulation and real robotic soccer leagues, such as the Sony Aibo in the Four-legged League and the Aldebaran Nao robots in the Standard Platform League.

In conclusion, a complex system of rewards is required to appraise the outcomes of a research program like RoboCup. This system takes into account both regional and general interests in academic research, double allegiance of participating scientists, both scientific and technological advances in machine intelligence, in addition to their translational promises in the way of industrial applications. Finally, one should not neglect a wide range of ancillary benefits depending on the allure of machine intelligence games.

These benefits range from the pedagogic payoffs of attracting wider cohorts of science and engineering students, to motivating effects for junior scientists, and to the outreach of dissemination actions using soccer as an expedient to engage audiences that are not familiar with robotic systems and technologies. An understanding of this composite and intricate system of rewards is also needed on the part of individual scientists who have to engage into multifaceted forecast, outcome appraisal, and timely deliberation processes in order to decide whether to confirm or to waive their secondary allegiance to RoboCup and other multidisciplinary research programs in machine intelligence.

5. Broader philosophical reflections on machine intelligence games

Gamification is a recently coined word referring to the application of game design principles in non-gaming contexts (Robson *et al*., 2015). Gamification appears to be an increasingly popular strategy in the definition and pursuit of research programs in machine intelligence. RoboCup initial focus on soccer playing has now expanded towards socially useful robotic tasks and related competitions between robotic systems. The long-term goal of RoboCup@home (www.robocupathome.org) is to develop autonomous robots which share with humans the physical space of their homes and carry out there a variety of household and assistive tasks. RoboCupRescue (www.robocuprescue.org) is a competition recruiting and comparing the search and rescue capabilities of robotic teams working in earthquake-stricken and other disaster areas. Gamification of research activities is a strategy pursued by major research funding schemes too. The EU Horizon 2020 program calls, on page 70 of its 2016-17 ICT Work Programme, for the «organisation of robotic competitions to speed up the advance towards smarter robot, demonstrating progress in the field and raising the awareness of the general public towards intelligent robots.» (https://ec.europa.eu/programmes/horizon2020/en/draft-work-programmes-2016-17).

It was pointed out above that gamification of research programs helps one to address and hopefully to sidestep methodological problems afflicting many traditional research programs in machine intelligence by insisting on clear rules of behavior and task achievement conditions. However, the present analysis suggests that there is a trade-off between this expected methodological benefit of research gamification and a corresponding cost: methodological difficulties that one hopes to kick out of the machine intelligence door by means of clear playing rules and game winning criteria come back through the window of a fairly composite and intricate appraisal system for machine intelligence research programs based on games.

It is worth noting that major military research funding agencies pursue gamification of research programs in machine intelligence. Notably, the Defense Advanced Research Projects Agency (DARPA) of the US Department of Defense (DoD) promotes the Robotic Challenge (www.darpa.mil/Our_Work/TTO/Programs/DARPA_Robotics_Challenge.aspx), which is driven by the long-term goal of building autonomous robotic systems replacing human operators in dangerous task environments. The Robotic Challenge develops through periodic robotic tournaments for task-level autonomy and operational capabilities in hazardous trial arenas. Even though these arenas are not directly identifiable with warfare scenarios, novel technologies and systems that one demonstrates there possess many features of interest to R&D departments of military industries. Accordingly, the system of rewards for appraising research programs based on games and competitions must be enriched so as to include military application payoffs. By the same token, the deliberation processes that scientists must engage into in order to decide whether to confirm or to waive their allegiance to RoboCup and other research programs on machine intelligence games must take into account the potential dual use of their research products and related ethical issues.

A thorough examination of the ethical issues that must be taken into account by scientists participating in research programs on machine intelligence games goes beyond the scope of this article. However, it is quite evident that dual use payoffs and related ethical issues must be duly taken into account in order to set up a comprehensive appraisal process for machine intelligence research programs based on games. Similarly, the narrative about the payoffs of the US space program should be extended so as to include US reactions to earlier achievements of the USSR space program, related fears of losing supremacy in spaceflight capabilities and arms races, in addition to emerging ethical issues and expected payoffs in the way of military applications, such as more extended ranges and heavier warhead payloads for US military missiles. Similarly, one cannot ignore that team coordination strategies used in RoboCup are potentially interesting for coordinating the action of uninhabited air, land, sea, or underwater vehicles for military applications in warfare.

Visionary goals for machine intelligence games have not been invariably productive of fruitful scientific and technological research programs. A case in point is Turing's imitation game, also known as Turing Test, and the related ambitious objective of a computer doing well in the game. The imitation game involves a human being, a computing machine, and another human being in the role of interrogator who communicates with the other two players by texting only. By asking them questions, the interrogator must discover which one of her interlocutors is the machine. A machine playing well the imitation game will bring its interrogators to make the wrong identification often enough after extended periods of conversational exchanges (Turing, 1950; Cordeschi 2002, pp. 171-172; see also Frixione, 2016). The natural language processing capabilities that are required to play well the imitation game are a potentially very rich problem-solving domain for machine intelligence inquiry. However, accomplishment criteria for the imitation game, unlike those for chess or soccer, are rather vague and open to ample subjective disagreement. Moreover, the long-term goal of passing Turing's test – whatever this may precisely mean – has not given rise to a series of feasible subgoals enabling one to advance computer playing skills through competition and ranking.

It is doubtful that Turing meant the imitation game as regulative idea for a genuine research program on machine intelligence games. Arguably, Turing's intentions were more pedagogic in character: back in 1950, when a handful of main frame computers were actually operating, the imitation game enabled him to convey vividly to the general public the idea that computing machines would gradually become able to compete and cooperate with humans in a wide variety of symbol-processing activities (Gandy, 1986, p. 125; Cordeschi and Tamburrini 2015). Be as it may, it is more interesting in the context of the present methodological discussion to emphasize that Turing addressed the question "Can machines think?" in the opening paragraph of his 1950 article on the imitation game, but immediately suggested to sidestep the problem of defining what is thinking and to focus instead on manifestations of biological and artificial intelligence occurring in game playing competitions. And in a later article Turing actively suggested to use chess and other social games as generators of research programs in machine intelligence (Turing, 1953). In particular, he was careful to distinguish between different levels of chess playing skills, thereby suggesting a series of subgoals for computer chess inquiries and the ranking of chess programs (Turing, 1953; see Turing, 1992, pp. 163-64).

It was recalled above that Alexander Kronrod, a Russian mathematician and computer scientist, stated in 1965 that chess is to AI what the model organism *Drosophila* is to biology. Here, computer chess and robotic soccer have been rather used as a *Drosophila* for the methodology of AI and robotics, insofar as a reflection on research programs that are based on machine game playing opens a window on central features of the complex systems of rewards and appraisal criteria that come into play in multidisciplinary machine intelligence investigations. In particular, these appraisal systems must take into account what has been called here the double allegiance of scientists, and the benefits for their primary allegiance flowing from work done on a multidisciplinary research program that scientists espouse as their secondary allegiance only. The sharing, standardization and trading benefits flowing from multidisciplinary research programs should not be belittled in a research area where intersubjective criteria to identify what counts as a result and properly weigh its importance are hard to come. These various ingredients should be properly taken into account and weighed to develop more accurate models of systems of rewards for scientists (Strevens, 2003; Weisberg and Muldoon, 2009; McKenzie Alexander, Himmelreich and Thomson, 2015) which aim at explaining observed patterns in the division of epistemic labor and to identify promising strategies in the context of scientific discovery.

References

Asada M., Kitano H., Noda I., Veloso M. (1999). RoboCup: today and tomorrow – what we have learned. *Artificial Intelligence* 110:193–214.

Cordeschi R. (1984). La teoria dell'elaborazione umana dell'informazione. Aspetti critici e problemi metodologici. In: Continenza B., Cordeschi R., Gagliasso E., Ludovico A. and Stanzione M., *Evoluzione e modelli. Il concetto di adattamento nelle teorie dei sistemi biologici, culturali e artificiali*. Roma: Editori Riuniti: 321-422.

Cordeschi R. (2002). *The Discovery of the Artificial. Behavior, Mind and Machines Before and Beyond Cybernetics,* Dordrecht: Kluwer. Doi: 10.1007/978-94-015-9870-5

Cordeschi R. and Tamburrini G. (2015). Alan Turing e il programma di ricerca dell'Intelligenza Artificiale. In Hosni H, a cura di, *Menti e macchine. Alan Mathison Turing a cento anni dalla nascita*. Pisa: Edizioni della Normale: 85-124.

Frixione M. (2016). The Turing Test and the interface problem: a role for the imitation game in the methodology of cognitive science. *Paradigmi,* this issue.

Gabel T. and Riedmiller M. (2010). On progress in RoboCup: the simulation league showcase. In: Chown E. and Plöger P.G., eds. *RoboCup 2010: Robot Soccer World Cup XIV, LNCS* 6556. Singapore: Springer.

Galison P. (1997). *Image and logic: a material culture of microphysics*. Chicago: University of Chicago Press.

Gandy R. O. (1996). Human versus mechanical intelligence. In: Millican P. and Clark A., eds, *Machines and Thought*. Oxford: Clarendon Press: 125-136.

Godfrey-Smith P. (2003) *Theory and reality: an introduction to the philosophy of science*. Chicago: University of Chicago Press.

Kitano H., Kuniyoshi Y., Noda I, Asada M., Matsubara H. and Osawa E. (1997) RoboCup: a challenge problem for AI. *AI Magazine* 18:73–85

Kuhn Th. S. (1962). *The structure of scientific revolutions*Chicago: University of Chicago Press. Fourth 50th anniversary edition (2012) *ibidem*.

Lakatos I. (1970) Falsification and the methodology of scientific research programmes. In: Lakatos I. and Musgrave A., eds, Criticism and the growth of knowledge, Cambridge University Press: 91-196. Reprinted in: Worrall J. and Currie G. eds, *Philosophical papers, vol. 1: the methodology of scientific research programmes*. Cambridge University Press.

Laudan L (1977) *Progress and its problems: toward a theory of scientific growth*. University of California Press, Berkeley

Levinson R., Feng-hsiung Hsu, Marsland T. A., Schaeffer J. and Wilkins D.E. (1991). Panel on Computer chess. In: Mylopoulos J. and Reiter R., eds, *Proceedings of the 12th International Joint Conference on Artificial Intelligence*. Sydney, Australia, August 24-30, 1991. San Francisco: Morgan Kaufmann: 547-552.

McCarthy J. (1997). AI as sport. *Science* 276: 1518-1519.

McKenzie Alexander J, Himmelreich J and Thomson Ch. (2015). Epistemic landscapes, optimal search, and the division of cognitive labor. *Philosophy of Science* 82: 424-453.

Robson K., Plangger K., Kietzmann J., McCarthy I., and Pitt L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons* 58: 411-420

Russell S. and Norvig P. (2010) *Artificial intelligence: a modern approach*. 3rd edition. New York: Prentice Hall,.

Strevens M. (2003). The role of the priority rule in science. *Journal of Philosophy* 100: 55-79.

Turing A.M. (1950). Computing Machinery and Intelligence. *Mind* 59: 433–460. Reprinted in (Turing 1992), pp. 133-160.

Turing, A. M. (1953), Digital computers applied to games. In Bowden B. V., ed. *Faster than Thought. A Symposium on Digital Computing Machines*. London: Pitman & Sons. Reprinted in (Turing 1992), pp. 161-185.

Turing A. M. (1992). *Collected Works of A. M. Turing*. Vol. 3, Ince D. C. ed. Amsterdam: North-Holland.

Visser U. and Burkhard H. D. (2007). RoboCup: ten years of achievements and future challenges. *AI Magazine* 28:115–132.

Weinsberg M. and Muldoon R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science* 76: 225-52.