

On the ethical framing of research programs in robotics

Guglielmo Tamburrini¹

Received: 11 November 2014 / Accepted: 6 October 2015 / Published online: 27 October 2015
© Springer-Verlag London 2015

Abstract Robotic systems and technologies step out of research laboratories jointly with information about long-term goals of technological inquiry they are lined up with and about the short-term objectives guiding daily laboratory activities. These various ingredients play crucial roles in the pursuit of what are called here *technological research programs*. A comprehensive ethical framing of technological research programs is decomposed here into the ethical framing of their long-term and short-term goals, respectively. This approach to the ethical framing of technological research is exemplified by reference to fundamental rights in the context of technological research programs on elderly care and child care robots. Moreover, its significance is highlighted in connection with democratic decision-making about new and emerging technologies, as well as in connection with the cultural production of ignorance which is induced by missing information about the protection and promotion of fundamental rights in the specific context of robotic technologies.

Keywords Applied ethics · Robotics · Technological research programs · Elderly care robots · Child care robots · Fundamental rights · Agnotology

1 Introduction

The invitation to reflect on ethical and societal challenges arising *beyond* the walls of robotic laboratories urges one to avoid the pitfalls of overly speculative approaches in applied ethics.¹ These speculative approaches tend to concentrate on novel ethical issues emerging in technologically distant scenarios that are not in the purview of current scientific and technological knowledge (Farah 2002; Nordmann 2007; Tamburrini 2009). In robotics, speculative drifts of ethical reflection are visible from discussions of, e.g., purported moral responsibilities of and toward robots, insofar as the construction of a robotic moral agent is neither a plausible working hypothesis of current research nor one of its foreseeable developments. One may effectively countervail speculative approaches in ethical reflection about robotics by focusing on actual inquiries that are carried out in robotics research laboratories and their expected outcomes. This recommendation, however, presupposes a relatively clear idea of what it is that one expects to step out of robotics laboratories. Let us then try and take a tentative stock.

Clearly, prototypical systems and technologies which promise to satisfy the mosaic of market viability requirements step out of research laboratories. These mature outcomes of technological inquiry are framed into research plans specifying both long-term objectives of inquiry and a series of short-term objectives to be reached

✉ Guglielmo Tamburrini
tamburrini@unina.it

¹ Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione (DIETI), Università di Napoli Federico II, Naples, Italy

¹ This paper is based on my presentation at the International Conference *Going beyond the laboratory: ethical and societal challenges for robotics* organized by Universität Oldenburg and held on February 13–15, 2014, at the Hanse Wissenschaftskolleg in Delmenhorst, Germany. My presentation there benefited from previous discussions of technological research programs with Roberto Cordeschi.

along the way. Objectives of both kinds are made known outside research laboratories to a variety of stakeholders, including research funding agencies, interested communities of scientists and engineers, political decision-makers, and the general public of citizens accessing media reports about science and technology. Accordingly, technologies step out of research laboratories jointly with—and on many occasions only after—the long-term goals of technological inquiry they are framed into, and the various short-term objectives more concretely guiding research laboratory activities.

This wide construal of what steps out of research laboratories is crucial for applied ethics, insofar as the formulation of long-term and short-term research goals selectively influences the paths of technological inquiry, affecting at the same time beliefs that citizens entertain about the human purposes that technology may or even ought to serve. According to this wide construal, ethical reflection on what goes beyond research laboratories should embrace each main component of *technological research programs* and their dynamic interrelations. By setting up a comprehensive ethical frame for technological research programs, one brings ethical reflection to bear on the various asynchronous activities taking place within and around research laboratories: from the formulation of ethically motivated visions of technological development, to the promotion of value-sensitive design and technology assessment, up to and including a proper understanding of ethical issues that are involved in public deliberations about new and emerging technologies.

The notion of technological research program (TRP from now on) is preliminarily examined in Sect. 2, and exemplified there by reference to RoboCup and other TRPs in robotics. The ethical framing of long-term goals of ambitious TRPs is discussed in Sect. 3, focusing on the ethical and legal frame of fundamental rights in the context of elderly care robots. Implications of this analysis for the short-term goals of TRPs are examined in Sect. 4 and exemplified there by reference to child care robots and, once again, elderly care robots. Finally, Sect. 5 examines the role of ethical framing in public deliberation processes about the development and use of new and emerging technologies, concentrating on the cultural production of ignorance induced by missing information about ethical frames for TRPs in robotics.

2 Technological research programs and robotics

Ambitious technological research programs in service and social robotics aim at developing robots that carry out useful tasks and interact closely with humans in homes, schools, offices, hospitals, entertainment environments, and

other typically human habitats. There, service and social robots are expected to fulfill roles of skilled helpers and tutors, dependable caregivers, trustworthy surveillance operators, dexterous assistants, or even enjoyable robotic companions. Many of these prospective applications of robotics are well beyond technological state of art and should be accordingly regarded as *long-term goals* of technological inquiry. These long-term goals guide the search for series of *short-term subgoals* that are expected to be feasible on the basis of present scientific and technological knowledge and are additionally expected to feed, with their technological pay-offs, the pipeline between research and industry. TRPs on elderly care robots, for example, are expected to release in the short term a variety of commercially viable robots for telepresence, eating support, medication taking reminders, and emergency assistance requests.

According to this construal, ambitious TRPs in robotics and other domains of technological investigation crucially involve the interaction of two asynchronous processes:

1. *Goaling* the process of identifying and updating long-term goals that are well beyond technological state of art;
2. *Subgoaling* the process of identifying, pursuing, and updating subgoals of long-term goals that are deemed to be both achievable in the short term and technologically rewarding on the basis of current scientific and technological knowledge.

Similar asynchronous processes are postulated in two-process models of *scientific inquiry* (Godfrey-Smith 2003), which notably include Lakatos's scientific research programs (Lakatos 1978) and Laudan's research traditions (Laudan 1977). Roughly speaking, one of the involved asynchronous processes is concerned with the development of a general framework, comprising a collection of ambitious scientific goals and a collection of theoretical ideas and techniques that are tentatively identified as appropriate means to pursue those ambitious scientific goals. The second asynchronous process is concerned with the identification and pursuit of a variety of short-term theorizing, modeling, and experimenting activities within the broad horizon afforded by the general framework.

A feedback loop is triggered by periodic appraisals of scientific results that have been obtained in the short term. The outcomes of these evaluations suggest whether to keep or to change short-term objectives, or even to modify the general framework in case of prolonged stagnation of research endeavors. Periodic appraisals of results obtained in the framework of *technological research*, as opposed to curiosity-driven scientific inquiry, crucially involve an evaluation of technological transfer to industry or other technological research areas.

The long-term goals of ambitious TRPs play a regulative role in technological inquiry. Expressing distant or even unattainable objectives for technological inquiry, long-term goals enable one to identify, in the present context of scientific and technological knowledge, subgoals that are feasible and are additionally expected to produce significant technological pay-offs.

To exemplify this broad description of a TRP and its asynchronous processes, let us consider more concretely the goaling and subgoaling processes that are involved in RoboCup. This is a TRP in robotics which is familiar to the general public for its robotic soccer tournaments, and its long-term goal of putting together a robotic soccer team which will eventually beat the human world champion team.

Fulfilling the long-term goal of a robotic soccer team defeating the human world champions presupposes so many far-reaching advances in sensorimotor and cognitive skills of multiagent robotic systems that one may sensibly doubt whether the research efforts of a few generations of committed scientists will suffice to bridge the gap between vision and reality. Therefore, the long-term goal of RoboCup is not a realistic objective to pursue for the current generation of robotic engineers. And it cannot serve either as a benchmark in periodic assessments of RoboCup contributions to technological progress, for robotic soccer teams are expected to perform very poorly against talented human teams for many years to come. This predicament is readily admitted in the RoboCup manifesto: 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.”² What is then the role, if any, of this elusive target in the context of RoboCup research activities?

The RoboCup long-term goal, it is surmised, enables one to shape a fruitful research agenda, insofar as it “can easily create a series of well-directed subgoals” that are both feasible and technologically rewarding. Indeed, winning teams of RoboCup tournaments enable one to identify a series of well-directed subgoals, insofar as their performances set new playing benchmarks that robotic teams participating in the next tournaments will be confronted with. Moreover, RoboCup steering committees identify more focused benchmarks concerning such skills as improved task-oriented feature selection or behavioral

pattern modeling in soccer playing multiagent environments. Finally, the outcomes of RoboCup subgoaling activities are selectively used to appraise the overall contributions of RoboCup to technological advancement with respect to prior states of art (Visser and Burkhard 2007; Gabel and Riedmiller 2010).

The founders of RoboCup emphasized from the very beginning the significance of entertaining an ambitious long-term goal for the purpose of developing a fruitful research agenda in the short term. In particular, they compared the role of the RoboCup visionary goal to the final goal of the US space program (Kitano et al. 1997; Asada et al. 1999), which President John F. Kennedy formulated, in his famous 1961 address to the US Congress, as the goal of “landing a man on the Moon and returning him safely to the Earth.” This ambitious goal was attained with the Apollo 11 space mission and culminated with the walk on the moon by astronauts Armstrong and Aldrin on July 20, 1969. During the lifecycle of this space program, more feasible technological subgoals were identified and pursued, in order to fulfill a wide variety of prerequisites for manned missions to the moon—from the identification of suitable fuel and missile technologies to medical interventions supporting human life in microgravity conditions. The achievement of short-term subgoals throughout the series of Apollo missions made a variety of new technologies available to industry and other research domains, thereby fertilizing many areas of technological inquiry and their industrial applications.

In addition to shaping a research agenda formed by a series of feasible and technologically rewarding subgoals, long-term goals play a distinctively *organizational* role in the pursuit of TRPs and the development of scientific communication and exchanges. Indeed, long-term goals contribute to establish what Peter Galison called *trading zones* (Galison 1997). These are virtual market places where investigators working in specialized scientific or technological domains come to exchange their models, technologies, and systems.³ Ambitious long-term goals of TRPs encourage cooperation and integration between research groups that usually pursue more regional objectives and do not interact with each other on a regular basis. Thus, for example, the long-term goal of RoboCup sets out “an integrated research task covering broad areas of AI and robotics. Such areas include: real-time sensor fusion,

² See <http://www.robocup.org/about-robocup/objective/>. This Webpage and every other Webpage referred to in this article were last visited on February 10, 2015. Here is the full quote: “Needless to say, the accomplishment of the ultimate goal will take decades of efforts, if not centuries. It is not feasible, with the current technologies, to accomplish this goal in any near term. However, this goal can easily create a series of well-directed subgoals. Such an approach is common in any ambitious, or overly ambitious, project.”

³ The wide range of application areas can be gleaned from *The Handbook of Robotics* (Siciliano and Khatib 2009) whose table of contents covers a wide variety of application domains (such as industrial and agricultural robotics, underwater, space and aerial robotics, search and rescue robotics, medical and rehabilitation robotics, and edutainment and social robotics) and activities that robots engage into (such as sensing and perceiving, moving, reasoning, and planning).

reactive behavior, strategy acquisition, learning, real-time planning, multiagent systems, context recognition, vision, strategic decision-making, motor control, intelligent robot control, and many more.”⁴

Periodic assessments of subgoal research activities activate TRP revision processes. Stagnating subgoal activities may induce one to revise the TRP short-term goals, and ultimately may even induce one to relinquish its long-term goal; sustained technological throughputs of subgoal activities buttress instead its long-term goal, and may even suggest the opportunity of extending research efforts toward more comprehensive long-term goals. Indeed, positive appraisals of RoboCup subgoal activities paved the way to RoboCup-Rescue and RoboCup@home.⁵

The long-term goal of RoboCup@home is to develop autonomous robots sharing with humans the physical space of their homes and carrying out there a variety of useful domestic and assistive tasks. This long-term goal displays—more clearly than the surface ambitions of RoboCup—deep-felt practical needs that often motivate research on new robotic systems and technologies. Let us turn now to examine, from the standpoint of normative ethics, some of the needs motivating RoboCup@home and the envisaged technological responses that are expected to come from current research efforts in assistive and social robotics.

3 Fundamental rights framing for elderly care robots

The aging population and a concomitant shortage of professional caregivers have been commonly adduced as practical motivations for long-term goals of TRPs on elderly care robots. Japan, whose aged population is presently the world’s highest, is expected to need *ceteris*

paribus 2.4 million nursing care workers in 2025, up from the 1.49 million employed in 2012 (*Japan Times*, 26 November 2013). This increased demand for specialized caregivers is likely to occur in a context of persistently low unemployment figures (3.5 % as of August 2014). As a proactive response to these upcoming needs, in November 2013 the Japanese Health, Labor and Welfare Ministry launched a campaign with the aim of promoting the use of robots in nursing homes for elderly care.

A broad moral conception of what society owes to dependent elderly humans underlies this governmental initiative (MacIntyre 1999). This moral conception is widely endorsed in human societies, but not invariably so, insofar as anthropological field observations document the practice of geronticide up to the present day (Diamond 2013, pp. 214–217). More important, one should be careful to note that in addition to robotic elder care there are other viable ways of meeting moral obligations toward dependent elderly humans. Indeed, there are both political and economic reasons suggesting reasonable alternatives to the robotic course of action envisaged in Japan and other parts of the industrialized world: the political decision to practice less restrictive immigration policies might alleviate expected shortages of specialized elder care personnel in Japan, and the lack of realistic estimates on the costs of still immature robotic elder care technologies may induce one to postpone their introduction until their economic viability will be convincingly supported by a comparative analysis of human and robotic care costs.⁶

Robotic solutions to elderly care have been questioned on more properly ethical grounds too. Philosophers Robert Sparrow and Linda Sparrow went as far as claiming that robotic elderly care should be expunged from the range of morally acceptable courses of action. Their argument is based on the premises that the social life of elderly people ought to be protected and that their often meager socialization levels will be further jeopardized by robotic caregivers: “It is likely that success in introducing robots into the aged-care sector will be at the expense of the amount of human engagement available to frail aged persons. We have highlighted the importance of social contact and both verbal and nonverbal communication to the welfare of older people. Any reduction of what is often already minimal human contact would, in our view, be indefensible” (Sparrow and Sparrow 2006, p. 152).

It is worth noting that the conclusion of this argument reflects the prevalent orientation of European citizens as of a 2012 statistical survey: EU citizens “... have well-defined views about the application areas for robots and the areas in which the use of robots should be banned: they

⁴ See again the Webpage <http://www.robocup.org/about-robocup/objective/>.

⁵ See www.robocupathome.org. RoboCup-Rescue is a friendly competition which aims at developing robotic systems working in multi-agent teams for search and rescue in earthquake-stricken and other disaster areas (www.robocuprescue.org). Competitive challenges have been a source of technological advancement in many different areas of technological inquiry. In robotics, another significant case in point is the DARPA Robotic Challenge, driven by the long-term goal of building autonomous robotic systems which enable one to replace human operators in dangerous task environments. The DARPA Robotic Challenge accommodates periodic competitions between robots for task-level autonomy and operational capabilities in a variety of hazardous trial arenas. See www.darpa.mil/Our_Work/TTO/Programs/DARPA_Robotics_Challenge.aspx. The automobile industry affords, with its various categories of formula racing competitions, another popular example of competition-driven technological research and innovation.

⁶ See IEEE Spectrum: <http://spectrum.ieee.org/automaton/robotics/home-robots/where-are-the-eldercare-robots/>.

should be used as a priority in areas that are too difficult or too dangerous for humans, like space exploration (52 % priority), manufacturing (50 %), military and security (41 %) and search and rescue tasks (41 %); there is widespread agreement that robots should be banned in the care of children, the elderly or the disabled (60 %).” (Special Eurobarometer Report “Public attitudes towards robots,” n. 382, September 2012). Similarly, negative attitudes emerged from a US survey: The prospect of robots replacing human workers in elder care drew a negative response among 65 % of Americans of all ages.⁷

In contrast with the motivations that Sparrow and Sparrow adduce for a ban on elderly care robots, some scientists who are actively involved in social robotics research argue that robots may improve the social life of elderly people. For example, Tony Prescott points out that “the provision of robot services such as telepresence can directly promote improved social interaction with friends and relatives for people who might have physical difficulties leaving the home” (Prescott 2013, p. 2). Similarly, a telepresence robot would enable someone who is confined home to participate actively in a museum guided tour or to enjoy the company of friends who are browsing around the shops of a shopping mall. Accordingly, if a ban on elderly care robots were put in place, one would miss significant opportunities for *promoting* the social life of elderly people.

Unsurprisingly, a combination of these different observations on robotic systems in elder care leads one to conclude that robotic systems—depending on what they are designed and used for—may in some cases jeopardize and in some other cases promote the social life of elderly people. Therefore, an indiscriminate ban on robotic elder care would lead one to throw out the baby with the bath water.

Interestingly, both of the above arguments are based on the shared premise that the social life of elderly people ought to be respected. Sparrow and Sparrow argue for a ban on the ground that the owed respect requires special protection from possible robotic disruptions; and Prescott argues against a ban emphasizing that robotic technologies may help one promoting this respect. In both cases, the crucial issue at stake is the respect of the social life of elderly people, which clearly involves complementary interventions for social life protection and promotion.

Complementary interventions concerning the respect of the social life of elderly people are morally grounded in various deontological and consequentialist theories in normative ethics through the intermediary of human autonomy conceptions and the relationships between autonomous action and social life (Mackenzie 2008).

Indeed, to be counted as an autonomous agent one must be able to act competently on the basis of one’s own desires and beliefs. In their turn, these enabling conditions for the exercise of human autonomy involve a self-interpretive capability, which comes into play in the process of identifying one’s own reasons for acting. But the capability for self-interpretation develops and is maintained through social interactions. Therefore, the respect of human autonomy—which is demanded by major deontological and consequentialist theories in normative ethics (Christman 2015)—entails the moral obligation to protect and promote the fragile preconditions for autonomous action that are found in the social life of elderly people.

The respect of personal autonomy and its preconditions in social life afford ethical motivations for more properly *legal* obligations to respect the social life of elderly people. Various international charters and treaties that are concerned with *fundamental* rights of human beings assert legally binding obligations to this effect. For example, articles 25 and 26 of the *EU Charter of Fundamental Rights* (EU 2000), which figure under the chapter heading *Equality*, assert the right of elder people to participate in social and cultural life, along with the right of disabled people to participate in the life of the community.

Fundamental rights have both proscriptive and prescriptive aspects. On the one hand, fundamental rights pertain to every human being and ought not to be violated. On the other hand, and on account of their prescriptive character, fundamental rights function as regulative ideas: their realization is an endlessly unfinished project, continually standing in need of contextual interpretation and assertion in the changing contexts of human life (Cohen and Arato 1992, p. 138; Camporesi 2011). Thus, in particular, fundamental rights must be reaffirmed and interpreted in connection with the growing impact of TRPs in robotics on the circumstances of human life. An adequate interpretation and assertion of fundamental rights in the context of robotic TRPs crucially involves a reflection on a wide variety of scientific models and empirical generalizations. The import of empirical knowledge on the ethical framing of TRPs in elderly care and child care robotics is selectively illustrated in the next section by reference to psychological implications of human–robot interactions (HRI), to attachment theory concerning mother–baby interactions, and to statistical correlations between impoverished social life and the insurgence of pathologies in elderly people.

4 Ethical framing and subgoaling

Fundamental rights must be reaffirmed and interpreted in the context of TRPs on elderly care robots, so as to identify subgoals that are both technologically feasible and coherent with fundamental rights protection and promotion.

⁷ See <http://spectrum.ieee.org/tech-talk/at-work/innovation/many-americans-wary-of-drones-robot-caregivers-and-google-glass/>.

On the protection side, the requirement that robots should not jeopardize the social life of elderly people suggests that one should refrain from pursuing, as a short-term goal, the development of robots which deceptively meet socialization needs by feigning intentionality, understanding, emotions, and empathy. Concerns about future scenarios of robots frustrating or falsely appeasing human communication needs have been variously expressed in connection with social robotics in general (Scheutz 2011) and with elderly care robotics in particular (Sharkey and Sharkey 2011). Present and foreseeable robotic systems are rather poor surrogates of human–human interactions. Therefore, they are likely bound to frustrate elementary needs for genuine verbal and nonverbal communication. In a more distant future, proficiently communicating robots will falsely appease human communication needs and will give rise to unidirectional projections of intentionality and understanding (unless, of course, they will achieve by then some kind of genuine intentionality and understanding).

On the promotion side, the right to social and community life inclusion is appropriately appealed to in order to support the development of robotic telepresence technologies enabling one to improve social interactions with relatives and friends. Additional ethical motivations for this short-term subgoal activity are found in the web of fundamental rights. For example, article 26 of the *Universal Declaration of Human Rights* (UN 1948) asserts that “everyone has the right to a standard of living adequate for the health and well-being of himself and of his family.” Medical investigations in the preconditions of good health and well-being enable one to relate meaningfully this human right to the development of telepresence robots. Indeed, it turns out that social integration prevents or delays the insurgence of various pathologies, thereby contributing to preserve the health and well-being of persons. When assessed by reference to a set of variables including the frequency of contact with children, parents, and neighbors, social integration was found to delay memory loss among elderly people (Ertel et al. 2008).

Respect of fundamental rights motivates the introduction of similar constraints or preferences in subgoal activities of TRPs on child care. According to article 19 of the *Convention on the Rights of the Child* (UN 1989), the child must be protected “from all forms of physical or mental violence, injury or abuse, neglect or negligent treatment, maltreatment or exploitation.” Forms of child neglect and maltreatment include chronically inadequate responsiveness by primary caregivers, who fail to supply the emotional security and encouragement that are usually needed in cognitive and emotional developmental processes. Against the background afforded by the psychological theory of attachment (Bowlby 1969–1980), protecting a child from neglect involves

protecting the synchronized communication and engagement taking place between babies or young children and their primary caregivers. These exchanges are crucially needed to achieve the sense of a secure basis from which to explore the world and to develop into a sufficiently self-confident adult. Psychological findings supporting these implications of attachment theory (Ainsworth et al. 1978) have been strengthened by more recent findings and models in developmental neuroscience, according to which early emotional and social deprivation interferes with normal development of the brain synaptic network, leading to behavioral and cognitive deficits in adult life, in addition to acquired vulnerability to stress-related psychiatric disorders (Schore 2010, pp. 129–130).

Clearly, no present or even foreseeable robotic system is capable of the genuine and subtle forms of communication that are observed across human cultures (see Eibl-Eibesfeldt 1989, chapter 7.1) and that are prized for their developmental effects in attachment theory. Therefore, the protection of the child from neglect and maltreatment demands that no robotic nanny be designed for the role of a child’s primary caregiver. More stringently, Sharkey and Sharkey (2010) argue that robots should be excluded from taking over the “dirty” job of cleaning babies, insofar as much synchronized communication between baby and primary caregiver originates in this basic form of childcare. Moreover, they sensibly suggest that robots should be allowed neither to take on key parental roles by, for example, restraining children actions, nor to spend unlimited amounts of time with babies or young children.

Within the boundaries suggested by these ethically motivated constraints, child care robots might fulfill a variety of useful roles, including those of educational support systems or whistleblowers for dangerous situations a child may get into. Feil-Seifer and Mataric (2010, p. 208) remark that “current research in socially assistive robotics is leading away from scenarios where a robot is the sole caregiver of a child.” This empirical evaluation of current work in robotics does not diminish the normative significance of ethically motivated constraints on subgoal activities of robotic child care TRPs, insofar as these constraints provide both guidance for responsible technological inquiry and ethical criteria for evaluating its outcomes. In the same article, the authors correctly point out that leaving a child in the care of a robot for extended periods of time would constitute neglect “as much as if the parent left the child under the supervision of other insufficient surrogate care, such as a television, video game, or an unqualified human caretaker. This neglect situation is not specific to a robot” (Feil-Seifer and Mataric 2010, p. 211). However, the observation that a wide variety of situations are subsumed under the umbrella term of neglect should not prevent one from appreciating the rich

differentiation between forms of child neglect and their specific threats to child well-being and psychological development. Thus, in particular, Matthias Scheutz is careful to note on the basis of a variety of HRI studies that “humans anthropomorphize robots, project their own mentality onto them, and form what seem like deep emotional, yet unidirectional relationships with them” (Scheutz 2011, p. 211), thereby leading to unidirectional emotional dependencies which differ in important respects from those arising in the context of, say, cell phone or PDA use.

5 Agnotology and ethical framing

In democratic societies, citizens responsibly participating in public affairs have to contribute to the formation of societal orientations or, more stringently, of policies and regulations about the development and use of new and emerging technologies. Relevant background information for these deliberative processes is provided by the ethical framing of both long-term and short-term goals of TRPs. Omitting this background information—which crucially includes information about personal obligations and societal commitments descending from the respect of fundamental rights—gives rise to significant *agnotological* effects, manifesting themselves in decision-making processes about the development and use of new technologies.

The word “agnotology” was introduced to designate the cultural production of ignorance and its study, especially in connection with its implications for individual and collective decision-making (Proctor 2008). A paradigmatic example of cultural ignorance production is the tobacco industry’s active policy of raising doubts about the dangers of smoking. Agnotological effects come with processes of *selective choice* too. For example, public calls for TRPs eventually prize some proposals and their research themes, while pruning other proposals for the extension of scientific and technological knowledge. It is surmised here that the deprivation of relevant background information about ethical framing is another significant source of ignorance, which is culturally produced by *default reasoning*, insofar as the missing information is replaced by default assumptions affecting the course of decision-making processes.

Information available to citizens about the long-term goals of TRPs on robotic elder care usually omits detailed reference to the frame of fundamental rights and its ethical underpinnings, emphasizing the more generic moral obligation to provide for elderly care in the present context of a rapidly aging population. The above-mentioned article from the *Japan Times* of November 26, 2013, is a significant case in point. If no detailed information is available about more specific moral obligations flowing from fundamental rights, and about the political commitments of states subscribing

international treaties on fundamental rights, one may suppose by default that there is no legal protection from forms of HRI that will, for example, deceptively fulfill the needs of elderly people in the way of genuine communication and empathy. From this supposition, those who regard the provision of adequate social interactions in elder care as a high moral priority might be led to sacrifice comparatively minor advantages offered by elder care robots and to call for a wholesale ban of these systems.

This line of argument is weakened by knowledge of what one is entitled to on the basis of an ethical framing of TRPs on robotic elderly care. By bringing out the implications of fundamental rights, one is made aware of legal principles and provisions protecting the social life of elderly people. This information suggests that there are ways to reap the potential benefits of technological research on robotic elder care while seeking protection from its potential threats to the social life of elderly people. More generally, the ethical framing of TRPs plays a crucial role in deliberations about the development and use of new technologies—offsetting the influence of incorrect default assumptions and pointing to what citizens are entitled to demand in the way of fundamental rights respect from national and international policies regulating the development and use of technologies.

Another agnotological effect which influences public discussion about new robotic technologies takes its origin in the formulation of the long-term goals of ambitious TRPs and propagates with the inadvertent contribution of scientists.

Scientists may issue statements about some TRP they are working on from the standpoint of its long-term goal or else from the standpoint of its short-term goals. It is unnecessary to signal which standpoint one is speaking from when scientists are addressing restricted circles of experts and especially those who are working on the same TRP. Experts are in a position to tell whether their peers are talking about some long-term visionary goal or else about its feasible—and usually much more modest—short-term subgoals. Discussion of long-term goals is important and highly prized in research communities, insofar as the identification of suitable long-term goals paves the way to fruitful subgoaling activities. In public statements, however, scientists can no longer count on the shared background of tacit knowledge which shapes communication styles within scientific communities. In particular, they cannot count on an understanding of the different time scales that are involved in pursuing a TRP’s long-term and short-term goals, respectively (Datteri and Tamburrini 2013).

Public statements about aims of TRPs which fail to convey the distinction between the time frames of long-term and short-term goals have agnotological effects that are amplified at the hands of media reporting and eventually contribute to generate the so-called hype cycles about new and emerging technologies (Seidensticker 2006; van Lente et al.

2013). Hype cycles are initially characterized by rising expectations about the technological promises of some TRP. Exaggerated expectations, however, will ultimately damage a TRP's credibility. Notably, TRPs in artificial intelligence were hit by a series of hype cycles culminating in the "AI winter" of the early 1990s (Russell and Norvig 2010, p. 24). In view of extensive media coverage and widespread anthropomorphizing projections on robots, TRPs in robotics are arguably taking their risks too.

Exaggerated psychological expectations flowing from a failure to convey correct information on the time frames of the long-term and short-term goals of TRPs in robotics are analyzed on the basis of psychoanalytic models of magic thinking and primitive mentality by Scalzone and Tamburrini (2013). According to these models, primitive minds assign to magic thinking the controlling and protective functions that one usually assigns to technology. Moreover, the omnipotence of thoughts and the wish-fulfilling attitude of magic thinking are still endorsed by children in a normal stage of their development and continue to operate unconsciously as repressed mental contents in adult mental life. Idealized views of robots that are conveyed by long-term goals of TRPs stir up these mental contents. Thus, in particular, the idealized perceptual, reasoning, planning, and acting capabilities of robots that are envisaged in long-term goals of robotic TRPs stir up the narcissistic illusion of attaining, by means of robotic systems, unlimited action and mental processing capabilities. In the long run, these narcissistic projections are likely to wane, as one realizes that real robots produced by means of TRP subgoal activities cannot meet unreasonable omnipotence expectations. As a consequence of this withdrawal of narcissistic projections from robots, the psychologically divested long-term goals of robotic TRPs fall into cultural devaluation. This psychoanalytic model contributes to account for *recurring* waves of technological hype as a matter of psychological need: newly developed TRPs supply fresh materials for omnipotence dreams that one cannot project any longer on their exhausted predecessors.

Correct information about the different temporal frames characterizing TRP long-term and short-term goals mitigates this psychologically rooted agnotological effect and its implications on the ethical issue of trust building between groups of stakeholders involved in the development and use of new technologies. Indeed, trust building between scientists and other stakeholders requires the development of suitable communication strategies on the part of scientists, enabling one to appreciate the achievement of TRP short-term goals and to set these tangible results apart from the formidable scientific and technological challenges that have to be met in the future in order to reach their admittedly visionary long-term goals.

6 Concluding remarks

The invitation to reflect on ethical and societal challenges for robotics arising *beyond* the walls of robotic laboratories requires a relatively clear idea of what it is that crosses the threshold of research laboratories. According to the wide construal proposed in this paper, the long-term goals of TRPs in robotics step outside research laboratories concomitantly with or ahead of many concrete robotic systems and technologies. These long-term goals selectively influence short-term paths of technological inquiry and affect at the same time beliefs about human purposes that technology can and ought to serve. For this reason, discussion of TRP long-term goals must be ethically framed so as to constrain TRP subgoal activities and to guide public debates and democratic decision-making processes about new and emerging technologies.

According to the view presented here, visionary technological scenarios that are not productive of feasible short-term subgoals hardly play any role in technological inquiry and related decision-making processes. Until the early 1960s, the idea of traveling to the moon was a visionary scenario playing no role in technological inquiry: Johannes Kepler's posthumously published work *Somnium* and Jules Verne's novel *From the Earth to the Moon* remind one that this visionary goal has been for centuries with the humankind without generating a genuine technological research program. Closer to robotics, one can point to a variety of visionary goals which, up to the present state of scientific and technological knowledge, have failed to play the generating role assigned here to the long-term goals of TRPs. These include visions of robotic systems that come to deserve the status of moral agents and the attendant attribution of fundamental rights, of robots that are conscious in the sense of having genuine subjective experiences, of robots serving as material substrates for the "download" of human minds. For all we know, these visions fall outside the scope of ethical reflection applied to robotic systems and technologies. The reason is not because these visions are too far-fetched: The RoboCup objective of putting together a robotic soccer team which will eventually beat the human world champion team is admittedly very remote and possibly unattainable. The rationale is rather that the RoboCup long-term goal is productively inserted—while these visions are not—into a set of asynchronous processes, dubbed here as technological research programs, which shape technological inquiry and effectively foster technological advancements.

The present approach aims at mitigating complementary defects of overly prospective and retrospective approaches in applied ethics, by taking jointly into account the main asynchronous processes of TRPs, their different time scales

and interactions. Overly prospective approaches focus on technological visions having a poor basis in state-of-art technology and playing no generative role in successful TRPs. Overly retrospective approaches concentrate on concrete systems that are on the verge of commercialization, without paying sufficient attention to the web of ideas and plans which led to their design and implementation. These retrospective approaches resemble Hegelian owls of Minerva in the context of applied ethics: spreading their wings at dusk, they may successfully capture the ethical implications of technologies produced over the past day of technological work, but this achievement is paid at the exorbitant cost of relinquishing the guiding role of ethical reflection in human-centered technological design (Cooley 2009) and in collective decision-making processes about new technologies.

Acknowledgments Stimulating comments by Thomas Christaller, Edoardo Datteri, Michael Nagenborg, Leen Spruit, and an anonymous reviewer are gratefully acknowledged.

References

- Ainsworth MD, Blehar M, Waters E, Wall S (1978) Patterns of attachment: a psychological study of the strange situation. Lawrence Erlbaum Associates, Hillsdale
- Asada M, Kitano H, Noda I, Veloso M (1999) RoboCup: today and tomorrow—what we have learned. *Artif Intell* 110:193–214
- Bowlby J (1969–1980) Attachment and loss. Hogarth, London
- Camporesi F (2011) The European Parliament and the EU charter of fundamental rights. In: Di Federico G (ed) *The EU charter of fundamental rights: from declaration to binding instrument*. Springer, Dordrecht, pp 77–94
- Christman J (2015) Autonomy in moral and political philosophy. *Stanford encyclopedia of philosophy*. Retrieved on June 9th, 2015 from <http://plato.stanford.edu/entries/autonomy-moral/>
- Cohen J, Arato A (1992) Politics and the reconstruction of the civil society. In: Honneth A, McCarthy Th, Offe C, Wellmer A (eds) *Cultural-political interventions in the unfinished project of enlightenment*. MIT Press, Cambridge, pp 121–144
- Cooley M (2009) On human-machine symbiosis. In: Gill S (ed) *Cognition, communication, and interaction*. Springer, London, pp 457–485
- Datteri E, Tamburrini G (2013) Robotic weapons and democratic decision-making. In: Hilgendorf E, Günther J-P (eds) *Robotik und Gesetzgebung*. Nomos Verlag, Baden-Baden, pp 211–230
- Diamond J (2013) *The world until yesterday*. Penguin, London
- Eibl-Eibesfeldt I (1989) *Human ethology*. Aldine de Gruyter, New York
- Ertel KA, Glymour MM, Berkman LF (2008) Effects of social integration on preserving memory function in a nationally representative US elderly population. *Am J Public Health* 98:1215–1220
- European Union (2000) *The EU charter of fundamental rights*. Retrieved on June 9th, 2015 from http://ec.europa.eu/justice/fundamental-rights/charter/index_en.htm
- Farah M (2002) Emerging ethical issues in neuroscience. *Nat Neurosci* 5:1123–1129
- Feil-Seifer D, Mataric MJ (2010) Dry your eyes: examining the roles of robots for childcare applications. *Interact Stud* 11:208–213
- Gabel T, Riedmiller M (2010) On progress in RoboCup: the simulation league showcase. In: Chown E, Plöger PG (eds) *RoboCup 2010: Robot Soccer World Cup XIV*, LNCS. Springer, Singapore
- Galison P (1997) *Image and logic: a material culture of microphysics*. University of Chicago Press, Chicago
- Godfrey-Smith P (2003) *Theory and reality: an introduction to the philosophy of science*. University of Chicago Press, Chicago
- Kitano H, Kuniyoshi Y, Noda I, Asada M, Matsubara H, Osawa E (1997) RoboCup: a challenge problem for AI. *AI Mag* 18:73–85
- Lakatos I (1978) Falsification and the methodology of scientific research programmes. In: Worrall J, Currie G (eds) *Philosophical papers: the methodology of scientific research programmes*, vol 1. Cambridge University Press, Cambridge
- Laudan L (1977) *Progress and its problems: toward a theory of scientific growth*. University of California Press, Berkeley
- MacIntyre A (1999) *Dependent rational animals: why human beings need the virtues*. Open Court, Chicago
- Mackenzie C (2008) Relational autonomy, normative authority and perfectionism. *J Soc Philos* 39:512–533
- Nordmann A (2007) If and then: a critique of speculative nanoethics. *Nanoethics* 1:31–46
- Prescott TJ (2013) Sunny uplands or slippery slopes? The risks and benefits of using robots in care. In: UKRE workshop on robot ethics, 25 March 2013, Sheffield, UK. Retrieved on June 9th, 2015 from <http://www.abrg.group.shef.ac.uk/pubs/?year=2013>
- Proctor R (2008) A missing term to describe the cultural production of ignorance (and its study). In: Proctor R, Schiebinger L (eds) *Agnology: the making and unmaking of ignorance*. Stanford University Press, Stanford, pp 1–33
- Russell S, Norvig P (2010) *Artificial intelligence: a modern approach*, 3rd edn. Prentice Hall, New York
- Scalzone F, Tamburrini G (2013) Human–robot interaction and psychoanalysis. *AI Soc* 28:297–307
- Scheutz M (2011) The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin P, Abney K, Bekey G (eds) *Robot ethics: the ethical and social implications of robotics*. MIT Press, Cambridge, pp 205–222
- Schore AN (2010) Attachment trauma and the developing right brain: origins of pathological dissociation. In: Dell PF, O’Neil JA (eds) *Dissociation and the dissociative disorders: DSM-V and beyond*. Routledge, New York, pp 107–143
- Seidensticker B (2006) *Future hype: the myths of technology change*. Berrett-Koehler Pub, San Francisco
- Sharkey A, Sharkey N (2010) The crying shame of robot nannies: an ethical appraisal. *Interact Stud* 11:161–190
- Sharkey A, Sharkey N (2011) Children, the elderly, and interactive robots. *IEEE Robot Autom Mag* 18(1):32–38
- Siciliano B, Khatib O (eds) (2009) *Handbook of robotics*. Springer, Berlin
- Sparrow R, Sparrow L (2006) In the hands of machines? The future of aged care. *Minds Mach* 16:141–161
- Tamburrini G (2009) Brain to computer communication: ethical perspectives on interaction models. *Neuroethics* 2:137–142
- United Nations (1948) *The universal declaration of human rights*. Retrieved on June 9th, 2015 from <http://www.un.org/en/documents/udhr/>
- United Nations (1989) *The convention on the rights of the child*. Retrieved on June 9th, 2015 from <http://www.unicef.org/UNICEFs-Work/UN-Convention/>
- van Lente H, Spitters C, Peine A (2013) Comparing technological hype cycles: towards a theory. *Technol Forecast Soc Change* 80:1615–1628
- Visser U, Burkhard H-D (2007) RoboCup: ten years of achievements and future challenges. *AI Mag* 28:115–132