Spring May 15, 2008

Safety Intelligence and Legal Machine Language: Do we need the Three Laws of Robotics?

Yueh-Hsuan Weng, *NCA, Ministry of the Interior, Republic of China*
Chien-Hsun Chen, *National Nano Device Laboratories*
Chuen-Tsai Sun, *National Chiao Tung University*

12

Safety Intelligence and Legal Machine Language: Do We Need the Three Laws of Robotics?

Yueh-Hsuan Weng, Chien-Hsun Chen and Chuen-Tsai Sun
Conscription Agency, Ministry of the Interior
National Nano Device Laboratories (NDL)
Dept. of Computer Science, National Chiao Tung University
Taiwan

1. Introduction

In this chapter we will describe a legal framework for Next Generation Robots (NGRs) that has safety as its central focus. The framework is offered in response to the current lack of clarity regarding robot safety guidelines, despite the development and impending release of tens of thousands of robots into workplaces and homes around the world. We also describe our proposal for a *safety intelligence* (SI) concept that addresses issues associated with open texture risk for robots that will have a relatively high level of autonomy in their interactions with humans. Whereas Isaac Asimov's Three Laws of Robotics are frequently held up as a suitable foundation for creating an artificial moral agency for ensuring robot safety, here we will explain our skepticism that a model based on those laws is sufficient for that purpose. In its place we will recommend an alternative *legal machine language* (LML) model that uses non-verbal information from robot sensors and actuators to protect both humans and robots. To implement a LML model, robotists must design a biomorphic nerve reflex system, and legal scholars must define safety content for robots that have limited "self-awareness."

2. Service robots

Since the Japanese already show signs of a special obsession with robots, it is no surprise that many new ideas on robot regulation are also emerging from Japan. Regarded as a "robot kingdom," it will most likely be the first country to produce and sell large numbers of NGRs for private use (Ministry of Economy, Trade and Industry [METI], 2004). That day is expected to emerge within the next two decades, raising both expectations and concerns among safety-conscious Japanese (Cabinet Office, Government of Japan [COGJ], 2007). Issued in February 2004, the Fukuoka World Robot Declaration contains details on Japanese expectations for emerging NGRs that will co-exist with and assist human beings, physically and psychologically. The guiding principle behind the document is to contribute to the realization of a safe and peaceful society (European Robotics Research Network [EURON], 2006); however, it fails to describe what NGRs should be.

In a report predicting the near future (2020-2025) in robot development, the Japanese Robot Policy Committee (RPC, established by METI) discusses two NGR categories: (a) next generation industrial robots capable of manufacturing a wide range of products in variable batch sizes, performing multiple tasks, and (unlike their general industrial predecessors) working with or near human employees; and (b) service robots capable of performing such tasks as house cleaning, security, nursing, life support, and entertainment—all functions that will be performed in co-existence with humans in businesses and homes (METI, 2004). The report authors predict that humans will gradually give NGRs a growing number of repetitive and dangerous service tasks, resulting in increased potential for unpredictable and dangerous actions. In a separate report published the following year, the RPC distinguished between the two NGR categories by listing three unique characteristics of service robots: strong mobility, close physical proximity to humans, and fewer repetitive operations (METI, 2005).

The Japanese Robot Association (JARA; http://www.jara.jp) predicts that next generation robots (NGRs) will generate up to ¥7.2 trillion (approximately $64.8 billion US) of economic activity by 2025, with ¥4.8 trillion going to production and sales and 2.4 trillion to applications and support (JARA, 2001). The Association divides production and sales into three domains: personal living (¥3.3 trillion), medical and social welfare (¥900 billion), and public service (¥500 billion). In summary, the NGR industry is currently emerging, and we can expect NGRs to enter human living spaces and start serving human needs within twenty years, making it imperative that we address safety and control issues now.

3. Robot safety

3.1 Robot sociability problems

Since 2000, Japanese and South Korean technocrats have been discussing and preparing for a human-robot co-existence society that they believe will emerge by 2030 (COGJ, 2007; Lovgren, 2006). Based on the content of policy papers and analyses published by both governments, researchers are currently studying potential robot sociability problems (RSP) that—unlike technical problems associated with design and manufacturing—entail robot-related impacts on human interactions in terms of regulations, ethics, and environments. Regulators are assuming that within the next two decades, robots will be capable of adapting to complex and unstructured environments and interacting with humans to assist with daily life tasks. Unlike heavily regulated industrial robots that toil in isolated settings, NGRs will have relative autonomy, thus allowing for sophisticated interactions with humans. That autonomy raises a number of safety issues that are the focus of this article.

3.2 NGR safety overview

In 1981, a 37-year-old factory worker named Kenji Urada entered a restricted safety zone at a Kawasaki manufacturing plant to perform some maintenance on a robot. In his haste, he failed to completely shut down the unit. The robot's powerful hydraulic arm pushed the engineer into some adjacent machinery, thus making Urada the first recorded victim to die at the hands of a robot ("Trust Me," 2006). This example clearly supports Morita et al.'s (1998) observation that when task-performing robots and humans share the same physical space, the overriding goal must be to ensure human safety. They note that several safety principles have already been adopted for industrial robots—for example, separation of operation space, fail-safe design, and emergency stop buttons. However, when NGRs and humans share the same physical space, it will be necessary to give robots the capabilities to protect biological beings and intelligence to use them.

Toward that end, METI invited experts from Japan's service and industrial robotics industries, lawyers and legal scholars, and insurance company representatives to participate in discussions of NGR safety issues. In July 2007 they published *Guidelines to Secure the Safe Performance of Next Generation Robots* (METI, 2007). According to Yamada (2007), the new guidelines are similar to those previously established for industrial robots in terms of risk assessment and risk reduction, but also contain five additional principles that make them unique to NGR safety requirements:

- In addition to manufacturers and users, parties who need to be involved in making decisions about NGR-related issues include company managers and sellers. In certain service domains, pedestrians and patients (in other words, those who will have the most contact with NGRs) should be included under the umbrella of NGR risk assessment.
- Accomplishing the goals of risk reduction requires multiple processing and testing procedures.
- Manufacturers need to cooperate with users, sellers, and managers when designing and making robots, especially during the risk assessment and reduction stages.
- Risk assessment procedures, safety strategies, and safety measures need to be clearly and thoroughly documented.
- Managers and sellers are obligated to notify manufacturers about accidents and to provide complete records related to those accidents for use in improving robot safety.

According to the METI guidelines, NGR manufacturers are obligated to enforce risk assessment rules and procedures during production. However, Kimura (2007) notes two major challenges to service robot risk assessment: their *openness* makes it difficult to clearly define users and task environments, and their *novelty* makes it hard to predict and calculate risk. Here we will add a third challenge: industrial robot safety involves machine standards, while autonomous NGR safety involves a mix of machine standards and *open texture risk* resulting from unpredictable interactions in unstructured environments (Weng et al., 2007).

In a May 2006 paper on legislative issues pertaining to NGR safety, Japanese METI committee members described the difference between industrial robots and NGRs in terms of *pre-* and *post-human-robot interaction responsibilities* (METI, 2006). In the following discussion we will refer to them as *pre-* and *post-safety regulations*.

For industrial robots, safety and reliability engineering decisions are guided by a combination of pre-safety (with a heavy emphasis on risk assessment) and post-safety regulations (focused on responsibility distribution). Pre-safety rules include safeguards regarding the use and maintenance of robot systems from the design stage (e.g., hazard identification, risk assessment) to the training of robot controllers. As an example of pre-safety rules, the United Kingdom Health and Safety Executive Office (2000) has published a set of guidelines for industrial robot safety during the installation, commissioning, testing, and programming stages. Another example is International Standardization Organization (ISO, 2006a) rules—especially ISO 10218-1:2006, which covers safety-associated design, protective measures, and industrial robot applications. In addition to describing basic hazards associated with robots, ISO rules are aimed at eliminating or adequately reducing risks associated with identified hazards. ISO 10218-1:2006 spells out safety design guidelines (e.g., clearance requirements) that extend ISO rules covering general machine safety to industrial robot environments (ISO, 2006b). Those rules thoroughly address safety issues related to control systems and software design, but since their primary focus is on robot arms and manipulators (ISO, 1994), they have limited application to NGRs.

3.3 Human based intelligence and open-texture risk

Neurologists view the human brain as having three layers (primitive, paleopallium, and neopallium) that operate like "three interconnected biological computers, [each] with its own special intelligence, its own subjectivity, its own sense of time and space, and its own memory" (MacLean, 1973). From an AI viewpoint, the biomorphic equivalents of the three layers are *action intelligence*, *autonomous intelligence*, and *human-based intelligence* (Fig. 1).

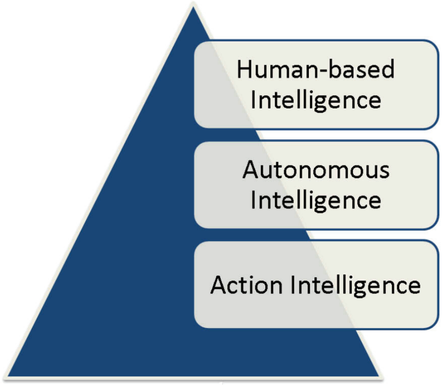

Figure 1. Robot Intelligence Layers.

Action intelligence functions are analogous to nervous system responses that coordinate sensory and behavioral information, thereby giving robots the ability to control head and eye movement (Hager et al., 1995), move spatially (Lewis, 1992), operate machine arms to manipulate objects (Chapin et al., 1999), and visually inspect their immediate environments (Dickmanns, 1988). Autonomous intelligence refers to capabilities for solving problems involving pattern recognition, automated scheduling, and planning based on prior experience (Koditschek, 1989). These behaviors are logical and programmable, but not conscious.

The field of robotics is currently in a developmental period that bridges action and autonomous intelligence, with robots such as ABIO (http://support.sony-europe.com/abio/), QRIO (Yeomans, 2005), and Roomba (http://www.irobot.com/sp.cfm?pageid=122/) on the verge of being lab tested, manufactured, and sold. These simple, small-scale robots are strong indicators of NGR potential and the coming human-robot co-existence age. Even as “pioneer” robots, they have remarkable abilities to perform specific tasks according to their built-in autonomous intelligence. For example, ABIO and QRIO robots have been programmed to serve as companions for the elderly, and Roomba robots are capable of performing housecleaning chores. However, none of them can make decisions concerning self-beneficial actions or distinguish between right and wrong based on a sense of their own value.

At the third level is human-based intelligence (HBI)—higher cognitive abilities that allow for new ways of looking at one's environment and for abstract thought; HBI is also referred to as "mind" and "real intelligence" (Neisser et al., 1996). Since a universally accepted definition of human intelligence has yet to emerge, there is little agreement on a definition for robot HBI. Many suggestions and predictions appear to borrow liberally from science fiction—for instance, Asimov's (1976) description of HBI robots forming a new species with the long-term potential of gaining power over humans. In real-world contexts, researchers are experimenting with ways of combining action, autonomous, and human-based intelligence to create robots that "comprehend complex ideas, learn quickly, and learn from experience" (Gottfredson, 1997).

HBI research started in the 1950s—roughly the same time as research on artificial intelligence (AI), with which HBI is closely associated. One of the earliest and most famous efforts at examining HBI potential entailed what is now known as the "Turing Test" (Turing, 1950). Taking a behaviorist approach, Turing defined human intelligence as the ability "to respond like a human being," especially regarding the use of natural language to communicate. There have been many efforts to create programs that allow robots to respond to words and other stimuli in the same manner as humans ("Emotion Robots," 2007), but no AI program has ever passed the Turing test and been accepted as a true example of HBI (Saygin et al., 2000). The legal and RSP issues that will arise over the next few decades are intricately linked with AI, which was originally conceived as "the science and engineering of making intelligent machines, especially intelligent computer programs" (McCarthy, 2007).

	Safety Design by Risk Assessment	**Safety Intelligence**
Risk	Machine Risk	Autonomous Behavior Risk (Open Texture Risk)
Limit	Machine Standards	Robot's Intelligence Architecture
Effect	Decrease Risk	Prevent dangerous behaviors

Table 1. A Comparison of Safety Regulation Methods.

Designed and constructed according to very specific standards, industrial robots are limited to performing tasks that can be reduced to their corresponding mechanisms—in other words, they cannot alter their mechanisms to meet the needs of changing environments. Therefore, the primary purpose for performing industrial robot risk assessments is to design mechanisms that match pre-approved safety levels (Table 1).

Complex NGR motions, multi-object interactions, and responses to shifts in environments resulting from complex interactions with humans cannot be reduced to simple performance parameters. Furthermore, residual risk will increase as robot intelligence evolves from action to autonomous to human-based, implying that current assessment methods will eventually lose their value. Instead, NGR designers/manufacturers and HBI developers will have to deal with unpredictable hazards associated with the legal concepts of *core meaning* and *open texture risk*. While all terms in a natural language have core meanings, the open texture characteristic of human language allows for interpretations that vary according to specific domains, points of view, time periods, and other factors—all of which can trigger uncertainty and vagueness in legal interpretations (Lyons, 1999). In addition to clearly defining core meanings, autonomous NGR designers and programmers must predict their acceptable and useful ranges.

In a policy paper published in May of 2007, the Japanese METI predicted that the addition of NGRs into human society will occur in three stages (METI, 2007). The focus of the first (current) stage is on "the diffusion of specific working object robots" to perform tasks such as cleaning, security, carrying documents or objects, and performing desk clerk duties. The second (expected to begin between now and 2010) will consist of "specific user and autonomously working object robots"—for example, nurse robots that can perform tasks to support the elderly. The final stage (beginning between 2010 and 2015) will primarily consist of multi-task autonomous ("universal") robots. METI describes the first stage as a "point" diffusion (meaning a low degree of human contact), the second as a "line" diffusion (medium degree), and the third as a "facet" diffusion (high degree). The extent to which each stage affects the daily lives of humans depends on the speed and content of AI development. Regardless of the specific characteristics of human-NGR interaction within each stage, open-texture risk will expand significantly with each transition. The inherent unpredictability of unstructured environments makes it virtually impossible that humans will ever create a fail-safe mechanism that allows autonomous robots to solve all open-texture problems.

4. Safety intelligence

4.1 Safety intelligence (SI) and Asimov's Three Laws.

NGR safety regulations will require a mix of pre- and post-safety mechanisms, the first entailing programming and using a robot's AI content to eliminate risk, and the second a product liability system to deal with accidents that do occur. A clear security issue will be limiting NGR "self-control" while still allowing them to perform designated tasks. As one *Roboethics Roadmap* author succinctly states, "Operators should be able to limit robot autonomy when the correct robot behavior is not guaranteed" (EURON, 2006).

Giving operators this capability requires what we will call *safety intelligence* (SI), a system of artificial intelligence restrictions whose sole purpose is to provide safety parameters when semi-autonomous robots perform tasks. Researchers have yet to agree on a foundation for a SI system, but the most frequently mentioned is Isaac Asimov's (1950) "Three Laws of Robotics," created for his science fiction novel, *I, Robot*:

- First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- Second Law: A robot must obey orders given it by human beings, except when such orders conflict with the First Law.
- Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

The first two laws represent a human-centered approach to SI that agrees with the current consensus of NGR designers and producers. As robots gradually take on greater numbers of labor-intensive and repetitive jobs outside of factories and workplaces, it will become increasingly important for laws and regulations to support SI as a "mechanism of human superiority" (Fodor, 1987). The third law straddles the line between human- and machinecentered approaches. Since the purpose of their functionality is to satisfy human needs, robots must be designed and built so as to protect themselves as human property, in contrast to biological organisms that protect themselves for their own existence. As one magazine columnist has jokingly suggested, "A robot will guard its own existence … because a robot is bloody expensive" (Langford, 2006).

In his introduction to *The Rest of the Robots* (1964), Asimov wrote, "There was just enough ambiguity in the Three Laws to provide the conflicts and uncertainties required for new stories, and, to my great relief, it seemed always to be possible to think up a new angle out of the 61 words of the Three Laws." While those ambiguities may be wonderful for writing fiction, they stand as significant roadblocks to establishing workable safety standards for complex NGRs. *Roboethics Roadmap* authors note that the Three Laws raise many questions about NGR programming: Which kinds of ethics are correct and who decides? Will roboethics really represent the characteristics of robots or the values of robot scientists? How far can and should we go when we program ethics into robots? (EURON, 2006)

Other robot researchers argue that Asimov's laws still belong to the realm of science fiction because they are not yet applicable. Hiroshi Ishiguru of Osaka University, the co-creator of two female androids named Repliee Q1 and Repliee Q2 (Whitehouse, 2005), believes it would be a mistake to accept Asimov's laws as the primary guiding principle for establishing robot ethics: "If we have a more intelligent vehicle [i.e., automobile], who takes responsibility when it has an accident? We can ask the same question of a robot. Robots do not have human-level intelligence" (Lovgren, 2007). Mark Tilden, the designer of a toy-like robot named RoboSapien, says "the problem is that giving robots morals is like teaching an ant to yodel. We're not there yet, and as many of Asimov's stories show, the conundrums robots and humans would face would result in more tragedy than utility" (ibid.). Ian Kerr (2007), law professor at the University of Ottawa, concurs that a code of ethics for robots is unnecessary:

> Leaving aside the thorny philosophical question of whether an AI could ever become a moral agent, it should be relatively obvious from their articulation that Asimov's laws are not ethical or legal guidelines for robots but rather about them. The laws are meant to constrain the people who build robots from exponentially increasing intelligence so that the machines remain destined to lives of friendly servitude. The pecking order is clear: robots serve people.

4.2 Three problems tied to Asimov's laws

Currently there are two competing perspectives on dealing with the mix of AI and safety: creating artificial agents with safety-oriented reasoning capabilities, or programming robots with as many rules as required for ensuring the highest level of safe behavior. Which perspective wins out will depend on how policy makers, designers, and manufacturers address three issues that we will address in the following subsections.

4.2.1 Legalities: safety intelligence

Future robot-related planning and design decisions will involve human values and social control, and addressing them will require input from legal scholars, social scientists, and public policy makers. A great amount of the data they will use to support their decisions must come from researchers familiar with robot legal studies. Levy (2006) argues convincingly that a new legal branch of robot law is required to deal with a technology that by the end of this century will be found in the majority of the world's households. Robot safety will be a major component of robot law. From an engineering viewpoint, SI entails self-control; from a legal viewpoint, SI entails automatic compliance and obedience to human rules. The need to add individual restrictions to autonomous robot behavior will be limited to those problems that pre-safety risk assessments cannot successfully address in order to meet two major goals of robot law: restricting human behaviors to prevent the misuse of robots, and restricting autonomous robot behaviors to mitigate open texture risk. Most likely a combination of these restrictions and carefully crafted legislation to let robots enforce the legal norm will be required to achieve a maximum state of robot safety.

As noted earlier, it is unrealistic to assume that SI will ever be able to eliminate all risk; instead, it is better to view SI as one part of a regulatory system that also contains a strong technological component. We believe the respective roles of designers, users, creators, and robots must be spelled out before the human-robot co-existence society emerges, while acknowledging that liability questions will require many court decisions to refine. For example, if Asimov's laws are used as a guideline, how will we enforce the idea that robots must not allow human beings to come to harm *through inaction*? Further, what about situations in which a NGR follows an incorrect human command?

4.2.2 Decision-making: do NGRs need doctrinal reasoning?

Giving autonomous robots the ability to define their own safety concepts means giving them the power to decide both when and how to react to stimuli. At some point such decisions will require artificial ethical and morality reasoning—the ability to distinguish between right and wrong. Robotists such as Shigeo Hirose (1989) argue that in conflicts involving doctrinal reasoning and morality, Asimov's Three Laws may become contradictory or be set aside in favor of poorly chosen human priorities. Using an extreme example, robots could be programmed to commit homicide under specific circumstances based on the wishes of a human majority. This example touches on two fears that many people have when they consider autonomous robots: they are troubled by the idea of letting robots obey rules that are difficult or impossible to express legislatively, and fearful of letting them defend laws established by imperfect humans.

The robot decision-making problem requires debate on both morality reasoning and legal content. Asimov himself acknowledged the potential for multiple contradictions between his Three Laws and doctrinal reasoning—many of them based on natural language, the medium used by humans to access legal content. Despite the many problems that human legal language present regarding vagueness and abstraction, laws and rules have the quality of being understood and followed via doctrinal reasoning. To argue that the same process can be used to accomplish SI means finding a universally accepted answer to the question, "Do NGRs need doctrinal reasoning to accomplish SI, and is it possible for them to acquire it?"

The focus of the new field of human-robot interaction (HRI) is communication and collaboration between people and robots. HRI researchers view human-robot communication as having verbal (using natural language) and non-verbal components (e.g., actions, gestures, non-linguistic sounds, and environment). Humans communicate using a mix of verbal and non-verbal information, with non-verbal information representing up to 60-70% of total communication (Kurogawa, 1994). We believe that non-verbal communication is a more reliable means for implementing SI and avoiding contradictions tied to doctrinal reasoning. The issue is necessarily complex because it entails AI, control engineering, human values, social control, and the processes by which laws are formed. In addition to being beyond the scope of this chapter, SI decisions are beyond the range of robotists' training and talent, meaning that they must work hand-in-hand with legal scholars.

4.2.3 Abstract thinking and the meaning of "safety"

The ability to think abstractly is uniquely human, and there is no way of being absolutely certain of how robots will interpret and react to abstract meanings and vague terms common to human communication. For example, humans know how to distinguish between blood resulting from a surgical operation and blood resulting from acts of violence. Making that distinction requires the ability to converse, to understand abstract expressions (especially metaphors), and to use domain knowledge to correctly interpret the meaning of a sentence. There are many examples that illustrate just how difficult this task is: Chomsky's (1957) famous sentence showing the inadequacy of logical grammar ("Colorless green ideas sleep furiously"), and Groucho Marx's line, "Time flies like an arrow, fruit flies like a banana" (http://www.quotationspage.com/quote/26.html). Such examples may explain why Asimov (1957) described robots as "logical but not reasonable."

If one day robots are given the power to think abstractly, then humans will have to deal with yet another legal problem: robo-rights. In 2006, the Horizon Scanning Centre (part of the United Kingdom's Office of Science and Innovation) published a white paper with predictions for scientific, technological, and health trends for the middle of this century ("Robots Could," 2006). The authors of the section entitled "Utopian Dream, or Rise of the Machines" raise the possibility of robots evolving to the degree that they eventually *ask* for special "robo-rights" (see also "Robot Rights?" 2006). In this paper we will limit our discussion to how current human legal systems (in which rights are closely tied to responsibilities) will be used when dealing with questions tied to early generations of non-industrial robots.

Whenever an accident involving humans occurs, the person or organization responsible for paying damages can range from individuals (for reasons of user error) to product manufacturers (for reasons of poor product design or quality). Rights and responsibilities will need to be spelled out for two types of NGRs. The system for the first type—NGRs lacking AI-based "self-awareness"—will be straightforward: 100% human-centered, in the same manner that dog owners must take responsibility for the actions of their pets. In other words, robots in this category will never be given human-like rights or rights as legal entities.

The second type consists of NGRs programmed with some degree of "self-awareness," and therefore capable of making autonomous decisions that can result in damage to persons or property. Nagenborg et al. (2007) argue that all robot responsibilities are actually human responsibilities, and that today's product developers and sellers must acknowledge this principle when designing first-generation robots for public consumption. They use two codes of ethics (one from the Institute of Electrical and Electronics Engineers and the other from the Association of Computing Machinery) to support their view that for complex machines such as robots, any attempt to remove product responsibility from developers, manufacturers, and users represents a serious break from human legal systems. We may see a day when certain classes of robots will be manufactured with built-in and retrievable "black boxes" to assist with the task of attributing fault when accidents occur, since in practice it will be difficult to attribute responsibility for damages caused by robots—especially those resulting from owner misuse. For this reason, Nagenborg et al. have proposed the following meta-regulation:

> If anybody or anything should suffer from damage that is caused by a robot that is capable of learning, there must be a demand that the burden of adducing evidence must be with the robot's keeper, who must prove her or his innocence; for example, somebody may be considered innocent who acted according to the producer's operation instructions. In this case it is the producer who needs to be held responsible for the damage.

If responsibility for robot actions ever reaches the point of being denied by humans, a major issue for legal systems will be determining punishment. In wondering whether human punishment can ever be applied to robots, Peter Asaro (2007) observes that :

> They do have bodies to kick, though it is not clear that kicking them would achieve the traditional goals of punishment. The various forms of corporal punishment presuppose additional desires and fears of being human that may not readily apply to robots—pain, freedom of movement, morality, etc. Thus, torture, imprisonment and destruction are not likely to be effective in achieving justice, reform and deterrence in robots. There may be a policy to destroy any robots that do harm, but as is the case with animals that harm people, it would be a preventative measure to avoid future harms rather than a true punishment. ... [American law] offers several ways of thinking about the distribution of responsibility in complex cases. Responsibility for a single event can be divided among several parties, with each party assigned a percentage of the total.

If we go this route, we may need to spell out robo-rights and responsibilities in the same manner that we do for such non-human entities as corporations; this is a core value in human legal systems. The question is whether or not we will be able to apply human-centered values to robots in the same manner. To practice "robot justice," those systems will be required to have separate sets of laws reflecting dual human/robot-centered values, and robot responsibilities would need to be clearly spelled out.

4.3 Three criteria and two critical issues for SI

As we have shown so far, Asimov's Three Laws of Robotics face several significant legal and engineering challenges, especially the need for collaboration among designers, AI programmers, and legal scholars to address robot sociability problems with safety as the guiding principle. Based on our belief that the importance of SI will increase as we rapidly approach a period in which NGRs are developed, tested, built, and sold on a large scale, we are currently working on a safety regulation model that emphasizes the role of SI during the pre-safety stage. In its current form, our proposal rests on three criteria for safe human-NGR interaction: (a) the ongoing capability to assess changing situations accurately and to correctly respond to complex real-world conditions; (b) immediate protective reactions in human-predictable ways so as to mitigate risks tied to language-based misunderstandings or unstable autonomous behaviors; and (c) an explicit interaction rule set and legal architecture that can be applied to all NGRs, one that accommodates the needs of a human-robot co-existence society in terms of simplicity and accountability. In order to apply the three criteria, there is a need to break down the safety intelligence concept into the two dimensions that are the focus of the next two sections. The first, which is ethics-centered, involves a special "third existence" status for robots or a complete ban on equipping NGRs with human based intelligence. The second involves the integration of third existence designation with a *legal machine language* designed to resolve issues associated with open texture risk.

5. Future considerations

In "ROBOT: Mere Machine to Transcendent Mind" (1999), Carnegie Mellon professor Hans Moravec predicts that robot intelligence will "evolve" from lizard-level in 2010, to mouse-level in 2020, to monkey level in 2030, and finally to human level in 2040—in other words, some robots will strongly resemble *first-existence* (biological) entities by mid-century. If his prediction is correct, determining the form and content of these emerging robots will require broad consensus on ethical issues in the same manner as nuclear physics, nanotechnology, and bioengineering. Creating consensus on these issues may require a model similar to that of the Human Genome Project for the study of Ethical, Legal and Social Issues (ELSI) sponsored by the US Department of Energy and National Institutes of Health (http://www.ornl.gov/sci/techresources/Human_Genome/research/elsi.shtml). Each agency has earmarked 3-5 percent of all financial support for genome research for addressing ethical issues. ELSI's counterpart across the Atlantic is the European Robotics Research Network (EURON), a private organization devoted to creating resources for and exchanging knowledge about robotics research (http://www.euron.org/). As part of its effort to create a systematic assessment procedure for ethical issues involving robotics research and development, EURON has published *Roboethics Roadmap* (2006), a collection of articles outlining potential research pathways and speculating on how each one might develop.

According to the *Roadmap* authors, most members of the robotics community express one of three attitudes toward the issue of roboethics:

- Not interested. They regard robotics as a technical field and don't believe they have a social or moral responsibility to monitor their work.
- Interested in short-term ethical questions. They acknowledge the possibility of "good" or "bad" robotics and respect the thinking behind implementing laws and considering the needs of special populations such as the elderly.
- Interested in long-term ethical concerns. They express concern for such issues as "digital divides" between world regions or age groups. These individuals are aware of the technology gap between industrialized and poor countries and the utility of developing robots for both types.

We (the authors of this paper) belong to the third category, believing that social and/or moral questions are bound to accompany the emergence of a human-robot co-existence society, and that such a society will emerge sooner than most people realize. Furthermore, we agree with the suggestions of several *Roboethics Roadmap* authors that resolving these ethical issues will require agreement in six areas:

- Are Asimov's Three Laws of Robotics usable as a guideline for establishing a code of roboethics?
- Should roboethics represent the ethics of robots or of robot scientists?
- How far can we go in terms of embodying ethics in robots?

- How contradictory are the goals of implementing roboethics and developing highly autonomous robots?
- Should we allow robots to exhibit "personalities"?
- Should we allow robots to express "emotions"?

This list does not include the obvious issue of what kinds of ethics are correct for robots. Regarding "artificial" (i.e., programmable) ethics, some *Roadmap* authors briefly touch on needs and possibilities associated with robot moral values and decisions, but generally shy away from major ethical questions. We consider this unfortunate, since the connection between artificial and human-centered ethics is so close as to make them very difficult to separate. The ambiguity of the term *artificial ethics* as used in the EURON report ignores two major concerns: how to program robots to obey a set of legal and ethical norms while retaining a high degree of autonomy (Type 1), and how to control robot-generated value systems or morality (Type 2). Since both will be created and installed by humans, boundaries between them will be exceptionally fluid.

5.1 "Learning" ethics

Two primary research categories in the field of artificial intelligence are *conventional* (or symbolic) and *computational*. Conventional AI, which entails rational logical reasoning based on a system of symbols representing human knowledge in a declarative form (Newell & Simon, 1995), has been used for such applications as chess games employing reasoning powers (Hsu et al., 1995), conversation programs using text mining procedures (http://www.alicebot.org), and expert systems that are used to organize domain-specific knowledge (Lederberg, 1987). While conventional AI is capable of limited reasoning, planning, and abstract thinking powers, researchers generally agree that the use of symbols does not represent "mindful" comprehension and is therefore extremely limited in terms of learning from experience (Dreyfus & Dreyfus, 1980).

Computational (non-symbol) AI (Engelbrecht, 2002) mimics natural learning methods such as genetic (Mitchell, 1996) or neural (Abdi, 1994). It supports learning and adaptation based on environmental information in the absence of explicit rules—an important first existence capability. Some advantages of computational AI are its capacities for overcoming noise problems, working with systems that are difficult to reduce to logical rules, and performing such tasks as robot arm control, walking on non-smooth surfaces, and pattern recognition. However, as proven by chess programs, computational AI is significantly weaker than conventional AI in abstract thinking and rule compliance. Still, many robotics and AI researchers believe that HBI in robots is inevitable following breakthroughs in computational AI (Warwick, 2004), while others argue that computational and conventional AI are both examples of behaviorism, and therefore will never capture the essence of HBI (Dreyfus & Dreyfus, 1980; Penrose, 1989). Those in the latter group claim that reaching such a goal will require a completely new framework for understanding intelligence (Hawkins & Blakeslee, 2006).

We believe the odds favor the eventual emergence of HBI robots, and therefore researchers must take them into consideration when predicting future robot safety and legal issues. They may agree with Shigeo Hirose of the Tokyo Institute of Technology that a prohibition on HBI is necessary (quoted in Tajika, 2001). Hirose is one of a growing number of researchers and robot designers resisting what is known as the "humanoid complex" trend, based on his strict adherence to the original goal of robotics: inventing useful tools for human use (quoted in Kamoshita, 2005). For Alan Mackworth, past president of the Association for the Advancement of Artificial Intelligence, the HBI issue is one of "should or shouldn't we" as opposed to "can or can't we" ("What's a Robot?" 2007). Arguing that goal-oriented robots do not require what humans refer to as "awareness," Mackworth challenges the idea that we need to create HBI for machines.

5.2 Third existence

In an earlier section we discussed the idea that robot responsibility should be regarded as human-owner responsibility. If we allow HBI robots to be manufactured and sold, the potential for any degree of robot self-awareness means dealing with issues such as punishment and a shift from human-centered to human-robot dual values. This is one of the most important reasons why we support a ban on installing HBI software in robots—perhaps permanently, but certainly until policy makers and robotists arrive at a generally accepted agreement on these issues. We also believe that creating Type 1 robots—in other words, programming robots to obey a set of legal and ethical norms while retaining a high degree of autonomy—requires agreement on human-centered ethics based on human values. The challenge is determining how to integrate human legal norms into robots so that they become central to robot behavior. The most worrisome issue is the potential of late-generation HBI robots with greater degrees of self-awareness to generate their own values and ethics—what we call Type 2 artificial ethics. Implementing Type 2 robot safety standards means addressing the uncertainties of machines capable of acting outside of human norms. We are nowhere near discussing—let alone implementing—policies for controlling HBI robot behavior, since we are very far from having HBI robots as part of our daily lives.

However, if the AI-HBI optimists are correct, the risks associated with HBI robots will necessitate very specific guidelines. A guiding principle for those guidelines may be the categorization of robots as "third existence," a concept created by Waseda University's Shuji Hashimoto (2003). Instead of living/biological (first existence) or non-living/non-biological (second existence), he describes third existence entities as machines that *resemble* living beings in appearance and behavior. We think this definition overlooks an important human-robot co-existence premise: most NGRs will be restricted to levels of autonomous intelligence that fall far short of HBI, therefore their similarities with humans will be minor. In addition, the current legal system emphasizes the status of robots as second-existence human property, which may be inadequate for those semi-autonomous NGRs that are about to make their appearance in people's homes and businesses, especially in terms of responsibility distribution in the case of accidents. In their analyses, Asaro (2007) suggests the creation of a new legal status for robots as "quasi persons" or "corporations," while Nugenborg (2007) emphasizes the point we made earlier about how robot owners must be held responsible for their robots' actions in the same way as pet owners.

6. Legal machine language

We believe there are two plausible strategies for integrating NGR safety intelligence with legal systems, one that uses natural languages and one that uses artificial machine languages. We will respectively refer to these as *legal natural language* (LNL) and *legal machine language* (LML) (Weng, et al., unpublished manuscript). The main goal of an LNL approach is to create machines capable of understanding ranges of commands and making intelligent decisions according to laws and rules written in a natural language. Since LNL is the primary medium for accessing legal content in human society, using it to program robots eliminates the need for an alternative medium, which is considered a difficult task requiring a combination of HBI and solutions to the structured-vs.-open texture issues discussed earlier. For the foreseeable future, NGRs will not be capable of making autonomous decisions based on a combination of legal natural language and an underlying principle such as Asimov's Three Laws.

Constructed languages (e.g., programming languages and code) can be used as bases for legal machine languages (LMLs) that express legal content. Two common examples already in use are *bots* (also called *agent software*) and service programs employed in agency network environments for providing content, managing resources, and probing for information. It is possible to use LML to control robot behavior—in Lawrence Lessig's (1999) words, "We can build, or architect, or code cyberspace to protect values that we believe are fundamental." We will give three examples of how LMLs are currently being used. First, the Internet Content Rating Association (ICRA) (http://www.fosi.org/icra) has created a standard content description system that combines features of a Resource Description Framework (RDF) and a Platform for Internet Content Selection (PICS). The system permits the marking of website content according to categories such as nudity, sex, or violence, and then allows parents and teachers to control online access. RDF makes it possible for machines to understand and react to ICRA labels embedded in websites as meta-tags. Whenever a browser with a built-in content filter (e.g., Microsoft Explorer's "Content Advisor") reads ICRA label data, it has the power to disobey user instructions based on "legal standards" established by parents or other authority. The second example is Robot Exclusion Standards (RES) (http://www.robotstxt.org/orig.html), also known as Robots Exclusion Protocol or robots.txt protocol. RES is currently being used to prevent web bots ("spiders") from accessing web pages that owners want to keep private, with the privacy policy written in a script or programming language. The third is Creative Commons (CC) (http://creativecommons.org/licenses/), used by copyright holders who want to control access in terms of sharing written information that they place on web sites. In addition to open content licensing, CC offers a RDF/XML metadata scheme that is readable by web bots. These examples show how humans can use constructed languages to communicate with robots in highly autonomous environments and to control their behavior according to agreed-upon legal standards.

Constructed languages are being examined as a means of overcoming the problem of ambiguity and the need to understand a range of commands associated with natural languages. Perhaps the best-known example is Loglan (http://www.loglan.org), identified as potentially suitable for human-computer communication due to its use of predicate logic, avoidance of syntactical ambiguity, and conciseness. The constructed language approach to programming commands in robots can be viewed as an intermediate step between human commands and autonomous robot behavior. As we stated above, NGRs in the foreseeable future will not be capable of making autonomous decisions based on a combination of legal natural language and an underlying principle such as Asimov's Three Laws. Since legal machine language does not give robots direct access to legal content, machines can be programmed to behave according to legal constraints without the addition of artificial ethics. Most likely a mix of LNL and LML will be required as NGRs interact more with humans, with human-machine communication based on simple rule sets designed to enhance safety when robots work around humans.

In an earlier section we discussed the neo-mammalian brain (HBI equivalent) functions of processing ethics, performing morality reasoning, and making right/wrong decisions. However, such a high level of reasoning is not required to recognize situations, avoid misunderstandings, and prevent accidents; a combination of action intelligence (unconscious) and autonomous intelligence (subconscious) is sufficient for these purposes. Nature gives us many examples of animals interacting safely without complex moral judgments or natural languages. When birds of the same species migrate, their shared genetic background allows them to fly in close or v-shaped formations without colliding—an important feature for collective and individual survival. This kind of safe interaction requires adherence to a set of simple non-verbal rules: avoid crowding neighboring birds, fly along the same heading, or at least along the same average heading as neighboring birds (Klein et al., 2003). The same flocking concept observed in birds, schools of fish, and swarms of insects is used to control unmanned aircraft and other machines (Gabbai, 2005; Reynolds, 1987). HBI-level moral and ethical decisions still require autonomous and action intelligence to correctly recognize environmental and situational factors prior to reacting in a safe manner.

6.1 Reflex control

Currently the focus of some exciting research, *reflex control* holds major potential as a tool allowing for the use of non-verbal communication and LML to command and control NGRs. Reflex control is a biomorphic concept based on the behaviorist principles of stimulus and response and the evolutionary principles of cross-generation inheritance. The basic function of reflexes, which are genetically built-in and learned by humans and animals, is to trigger protective behaviors within and across species, as well as between a species and its environment. Today we have a much clearer understanding of reflex action (and how it might be used to control robot behavior) than we do of complex intelligence. Some researchers are experimenting with sensors linked to reflexive controls—a combination known as "experiential expert systems" (Bekey & Tomovic, 1986)—for motion control (Zhang et al., 2002), automatic obstacle avoidance (Newman, 1989), collision avoidance (Wikman et al., 1993), and other simple behaviors associated with autonomous and action intelligence. Instead of requiring complex reasoning or learning capacities, robots equipped with reflex action can navigate fairly complex environments according to simple sets of common rules. Other robotists are working on a combination of reflex control and the ability to learn new protective behaviors for biomorphic robots, perhaps with a mechanism for transferring new knowledge to other machines. From a SI standpoint, reflex control allows for a high degree of safety in robot behavior because reactions can be limited to direct and immediate danger and programmed according to explicitly mapped responses. This meets our criteria for SI as a fast, clear, non-verbal, and predictable approach. Different sets of simple rules may be developed for NGRs operating in different environments (or countries) while still adhering to a "safety gene" standard (an explicit legal architecture applicable to all kinds of NGR robots), thereby satisfying another SI criteria.

An important distinction between human-designed robot reflex control and a NGR safety intelligence unit is that the SI unit would include content based on societal requirements to address open texture risk. Again, such content will require input from specialists to ensure compliance with accepted legal principles. If robotists ever reach a level of technological expertise that allows for HBI-programmed machines, a decision will have to be made about whether or not robots should be given human capabilities, or if reflex actions are sufficient for the needs and purposes of a human-robot co-existence society.

7. Conclusion

By the middle of this century, artificial intelligence and robot technology will no longer be considered science fiction fantasy. At the same time that engineers will be addressing all kinds of technical issues, a combination of engineers, social scientists, legal scholars, and policy makers will be making important decisions regarding robot sociability. In all cases, the topmost concern must be robot safety, since the emphasis for the future will be on human-robot *co-existence*.

In this paper we described a *safety intelligence* concept that meets three criteria: understanding situations, making decisions, and taking responsibility. The SI concept can also be broken down into two dimensions, the first involving special "third existence" ethical guidelines for robots plus a ban on equipping NGRs with human-based intelligence, and the second involving a mix of third existence designation and legal machine language designed to resolve issues associated with open texture risk. If current reflex control research proves successful, robots will someday be equipped with a safety reflex system to ensure that they avoid dangerous situations, especially those involving humans.

8. Acknowledgements

The authors would like to thank Mr. Jon Lindemann for his comments on the original manuscript.

9. References

Abdi, H. (1994). A neural network primer. *Journal of Biological Systems,* Vol. 2, 247-281.

Asaro, P. (2007). Robots and responsibility from a legal perspective. Paper presented at the *IEEE ICRA '07 Workshop on Roboethics*, Rome. Available online at http://www.roboethics.org/icra07/contributions/ASARO%20Legal%20Perspective.pdf.

Asimov, I. (1950). *I, Robot*. Doubleday: New York.

Asimov, I. (1957). *The Naked Sun*. Doubleday: New York.

Asimov, I. (1964). *The Rest of the Robots*. Doubleday: New York.

Asimov, I. (1976). *Bicentennial Man*. Ballantine Books: New York.

Robots Could Demand Legal Rights. (2006). *BBC News*, December 21. Available online at http://news.bbc.co.uk/2/hi/technology/6200005.stm.

Bekey, G. A. & Tomovic, R. (1986). Robot control by reflex actions. *Proceedings of the 1986 IEEE International Conference, Part 3: Robotics and Automation*, San Francisco, pp. 240-247.

Cabinet Office Government of Japan (COGJ). (2007). *Long-term strategic guidelines "Innovation 25"* (unofficial translation). Available online at http://www.cao.go.jp/innovation/index.html.

Chapin, J. K., Moxon, K. A., Markowitz, R. S. & Nicolelis, M. A. L. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neuroscience*, Vol. 2, 664-670.

Chomsky, N. (1957). *Syntactic Structures*. Mouton: The Hague/Paris.

Dickmanns, E. D. (1998). Dynamic computer vision for mobile robot control. *Proceedings of the 19th* International *Symposium Expos. Robots*, Sydney, Australia, pp. 314-327.

Dreyfus, H. L. & Dreyfus, S. E. (1980). *From Socrates to expert systems: The limits and dangers of calculative rationality*. Available online at http://Socrates.berkeley.edu/%7Ehdreyfus/html/paper_socrates.html.

Emotion Robots Learn from People. (2007). BBC News, February 23, available online at http://news.bbc.co.uk/2/hi/technology/6389105.stm.

Engelbrecht, A. P. (2002). *Computational intelligence: An introduction*. John Wiley & Sons: Hoboken, New Jersey.

European Robotics Research Network (EURON). (2006). *Roboethics Roadmap, Release 1.1*, Author: Genova, Italy.

Fodor, J. A. (1987). Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In Z. Pylyshyn (ed.), *The robot's dilemma: The frame problem in artificial intelligence*. Ablex Publishers: Norwood, New Jersey.

Gabbai, J. M. E. (2005). *Complexity and the aerospace industry: Understanding emergence by relating structure to performance using multi-agent systems*. Unpublished doctoral dissertation, University of Manchester.

Gottfredson, L. S. (1997). Mainstream science on intelligence. *Intelligence,* Vol. 24, Issue 1, 13–23.

Hager, G. D., Chang, W.-C. & Morse, A. S. (1995). Robot hand-eye coordination based on stereo vision. *IEEE Control Systems*, Vol.15, Issue 1, 30-39.

Hashimoto, S. & Yabuno, K. (2003). *The book of Wabot 2*. Chuko: Tokyo (in Japanese).

Hawkins, J. & Blakeslee, S. (2006). *On intelligence*. Yuan-Liou: Taipei (Chinese translation).

Hirose, S. (1989). A robot dialog. *Journal of Robotics Society of Japan*, Vol. 7, Issue 4, 121-126 (in Japanese).

Hsu, F. H., Campbell, M. & Hoane A. J. Jr. (1995). Deep Blue system overview, *Proceedings of the Ninth International Conference on Supercomputing*, Barcelona, pp. 240-244.

Huxley, A. L. (1932). *Brave New World*. Harper & Brothers: New York.

International Organization for Standardization (ISO). (1994). ISO 8373. Manipulating industrial robots – Vocabulary.

International Organization for Standardization (ISO). (2006a). ISO 10218. Robots for industrial environments - Safety requirements - Part 1: Robot.

International Organization for Standardization (ISO). (2006b). ISO 13849-1. Safety of machinery - Safety-related parts of control systems - Part 1: General principles for design.

Kamoshita, H. (2005). *The present and future of robots*. X-media: Tokyo (in Japanese).

Kerr, I. (2007). Minding for machines. *Ottawa Citizens News*. Available online at http://www.canada.com/ottawacitizen/news/opinion/story.html?id=e58202bb-f737-4ba7a0as-79e8071a1534.

Kimura, T. (2007). Risk assessment of service robots and related issues. *Journal of Robotics Society of Japan*, Vol. 25, Issue 8, 1151-1154 (in Japanese).

Koditschek, D. E. (1989). Robot planning and control via potential functions. *The Robotics Review,* Vol. 1, 349-367.

Kurogawa, T. (1994). *Nonverbal interface*. Ohmsha: Tokyo (in Japanese).

Langford, D. (2006). "It's the law" (magazine column). *SFX*, Issue 146. Available online at http://www.ansible.co.uk/sfx/sfx146.html

Lederberg, J. D. (1987). How Dendral was conceived and born. *Proceedings of the 1987 Association for Computing Machinery Conference on History of Medical Informatics,* Bethesda, Maryland, pp. 5-19.

Lessig, L. (1999). *Code and other laws of cyberspace.* Basic Books Press: New York.

Levy, D. (2006). *Robots unlimited: Life in a virtual age*. A. K. Peters: Wellesley, Massachusetts.

Lewis, M. A. (1992). Genetic programming approach to the construction of a neural network for control of a walking robot, *Proceedings of the IEEE International Conference on Robotics and Automation*, Nice, France, pp. 2118-2623.

Lovgren, S. (2006). A Robot in Every Home by 2020, South Korea Says, *National Geographic News* Sep. 6, 2006, available online at http://news.nationalgeographic.com/news/2006/09/060906-robots.html.

Lovgren, S. (2007). Robot Codes of Ethics to Prevent Android Abuse, Protect Humans, *National Geographic News,* March 16, 2007, available online at http://news.nationalgeographic.com/news/2007/03/070316-robot-ethics.html.

Lyons, D. (1999). Open texture and the possibility of legal interpretation. *Law and Philosophy*, Vol. 18, Issue 3, 297-309.

MacLean, P. D. (1973). A triune concept of the brain and behavior. In T. Boag & D. Campbell (eds.), *Hincks Memorial Lectures*, University of Toronto Press.

McCarthy, J. (*2007*). *What is Artificial Intelligence?* Available online at http://www-formal.stanford.edu/jmc/whatisai.html.

METI Robot Policy Council. (2005). *Robot policy midterm report.* Available online at http://www.meti.go.jp/policy/robotto/chukanhoukoku.pdf (in Japanese).

METI Robot Policy Council. (2006). *Robot policy report.* Available online at http://www.meti.go.jp/press/20060516002/robot-houkokusho-set.pdf (in Japanese).

Ministry of Economy, Trade and Industry (METI). (2004). *Toward 2025 and the human-robot co-existence society: The next generation robot vision seminar report*. Available online at http://www.meti.go.jp/kohosys/press/0005113/0/040402robot.pdf (in Japanese).

Ministry of Economy, Trade and Industry (METI). (2007). *Overview of METI robot policy.* Available online at http://www.meti.go.jp/policy/robotto/0705gaiyo.pdf (in Japanese).

Ministry of Economy, Trade and Industry (METI). (2007). *Guidelines to secure the safe performance of Next Generation Robots.* Available online at http://www.meti.go.jp/press/20070709003/02_guideline.pdf (in Japanese).

Mitchell, M. (1996). *Introduction to genetic algorithms*. MIT Press: Cambridge, Massachusetts.

Moravec, H. (*1999*). *ROBOT: Mere machine to transcendent mind.* Oxford University Press: New York.

Morita, T., Suzuki, Y., Kawasaki, T. & Sugano, S. (1998). Anticollision safety design and control *methodology* for human-symbiotic robot manipulator. *Journal of Robotics Society of Japan*, Vol. 16, Issue 1, 102-109 (in Japanese).

Nagenborg, M., Capurro, R., Weber, J. & Pingel, C. (2007). Ethical regulations on robotics in Europe. *AI & Society* (online edition), available online at http://www. springerlink. com/content/3t2m260588756t7v/ .

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, Vol. 51, 77–101.

Newell, A. & Simon, H. A. (1995). GPS: A program that simulates human thought. In Feigenbaum, E. A. & Feldman, J. (eds.) *Computers and Thought*. MIT Press: Cambridge, Massachusetts.

Newman, W. S. (1989). Automatic obstacle avoidance at high speeds via reflex control. *Proceedings of the 1989 IEEE International Conference, Part 2: Robotics and Automation*, Scottsdale, Arizona, pp. 1104-1109.

Penrose, R. (1989). *The emperor's new mind*. Oxford University Press: New York.

Saygin, A. P., Cicekli, Y. & Akman, V. (2000). Turing Test: 50 years later. *Minds and Machines*, Vol. 10, Issue 4, 463-518.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, Vol. 3, Issue 3, 417-457.

Spector, L., Klein, J., Perry, C. & Feinstein, M. (2005). Emergence of collective behavior in evolving populations of flying agents. *Genetic Programming and Evolveable Machines,* Vol. 6, Issue 1, 111-125.

Tajika, N. (2001). *The future Astro Boy*. ASCOM: Tokyo (in Japanese).

Robot Rights? It Could Happen, U.K. Government Told. (2006). *CBC News*, December 21, available online at http://www.cbc.ca/technology/story/2006/12/21/tech-freedom.html.

Trust me, I'm a robot. (2006). *The Economist*, June 8, available online at http://www. economist.com/displaystory.cfm?story_id=7001829.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, Vol. LIX, Issue 236, 433-460.

United Kingdom Health and Safety Executive Office. (2000). HSG 43:2000. Industrial robot safety.

Warwick, K. (2004). *March of the machines*. University of Illinois Press: Chicago.

Weng, Y. H., Chen, C. H. & Sun, C.T. (2007). The legal crisis of next generation robots: On safety intelligence. *Proceedings of the 11th International Conference on Artificial Intelligence and Law (ICAIL'07)*, Palo Alto, California, pp. 205-210.

Weng, Y. H., Chen, C. H. & Sun, C.T. (n. d.). Toward Human-Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots, *ExpressO (Unpublished Manuscript),* Available online at http://works.bepress.com/weng_yueh_hsuan/1.

What's a Robot? (2007). *CBC News*, July 16, http://www.cbc.ca/news/background/tech /robotics/definition.html.

Whitehouse, D. (2005). Japanese Develop "Female" Android. *BBC News,* July 27. Available online at http://news.bbc.co.uk/1/hi/sci/tech/4714135.stm.

Wikman, T. S., Branicky, M. S. & Newman, W. S. (1993). Reflexive collision avoidance: A generalized approach. *Proceedings of the 1993 IEEE International Conferenc*e, *Part 3: Robotics and Automation,* Kobe, Japan, pp. 31-36.

Yamada, Y. (2007). Currently existing international/domestic safety standards associated with service robots and ongoing tasks. *Journal of Robotics Society of Japan*, Vol. 25, Issue 8, 1176-1179 (in Japanese).

Yeomans, M. (2005). Qrio dances into spotlight at Carnegie Mellon Center. *Pittsburgh Tribune-Review*, January 29. Available online at http://www.pittsburghlive.com/x/pittsburghtrib/s_298114.html.

Zhang, X., Zheng, H., Duan, G. & Zhao, L. (2002). Bio-reflex-based robot adaptive motion controlling theory. *Proceedings of the 4th World Congress on Intelligent Control and Automation*, Vol. 3, pp. 2496-2499.