The Study of Interaction through the Development of Androids

Karl F. MacDorman[†,††] and Hiroshi Ishiguro[†,†††]

The development of androids is key to exploring human activity because only very humanlike machines can elicit natural responses in people. Conversely, insights drawn from detailed investigations of human activity are needed to build androids with humanlike behavior. We need to establish the new field of android science owing to the interdependent relationship between developing androids and investigating the mechanisms that support social activity. Certain questions about human beings cannot be answered without employing androids experimentally. Androids provide the ultimate test bed for theories from the social and brain sciences and a platform for their eventual unification.

1. Introduction

Androids present a vision of the future that most people first encountered in science fiction books, comics, or films. In *Terminator* they were cold and deadly, and in *A.I. Artificial Intelligence* at least one android exhibited a depth of feeling that rivaled its human counterparts. To develop androids that could truly mimic human beings poses challenges that far exceed the Apollo missions to the moon. Before taking on this challenge, it is important to consider why it might be justified.

We define an android to be an artificial system that has humanlike behavior and appearance and is capable of sustaining natural relationships with people. While we cannot know the potential benefits of developing androids in advance, devices that use appearance and behavior to establish relationships can be used to open and address a wide range of scientific questions. The value of building androids as opposed to mere humanoids is not yet widely accepted in the robotics community. Since the publication of *The Media Equation*,[32] it has become clear that human beings, having evolved in a social context, often treat computers, robots, and other media as if they were people. Why then do we need to build robots that really do look and act like people? This paper offers the beginnings of an answer.

† Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University
†† Frontier Research Center, Graduate School of Engineering, Osaka University
††† Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International

2. Observations from our androids

In our laboratory we have two androids, a replica of a five-year-old Japanese girl (**Fig. 1**) and a woman android, *Andosan*, with 31 degrees of freedom in the upper body, soon to be extended to 41 (**Fig. 3**). When Andosan is just idling (e.g., shifting posture, blinking, "looking" about), we often show visitors her reactive behaviors. Since she has been programmed to respond to touch sensors located on her head, shoulders, and arms, if we, for example, gently tap her shoulder, she will turn and ask "What is it?" in Japanese. However, if we slap her or hit her face, she winces, pulls back, and lifts her forearm to protect herself.

Guests typically cannot help but feel sympathy for Andosan in these moments, nor can they be enticed to treat her so rudely. This contrasts with our demonstrations of our humanoid robots' reactions to rough handling. People are inclined to feel far less compassion for Robovie or Eveliee, a robot based on Mitsubishi's Wakamaru platform. Their "robotic" appearance and behavior cannot elicit the same conscious and subconscious responses that the android does. And, indeed, children who are inclined to mob and play roughly with Robovie show deference and respect in Andosan's presence. Those of us who are developing Andosan are more willing to treat her as a mechanism just as surgeons can treat a person as a body. But people from outside our project are inclined to treat her as if she were a person, although they must surely know that she is not.

These examples show that to facilitate the most natural and humanlike interaction, we must build androids. Since most of our responses are nonconscious and inaccessible to

Fig. 1 The android *Repliee R1* is a replica of a five-year-old Japanese girl. Her arms, neck, head, eyes, and eyelids are powered by DC motors fixed inside the robot. Her skin is made of silicone. In the photograph on the left, the android's skin, hair, and padding have been removed.

introspection, simply "knowing" what a robot can do is not enough. The android form immediately tells us what the robot affords[13) –] or ought to! – and in a way that a merely humanoid form cannot. The fact that Andosan looks and is beginning to act like a capable Japanese woman sets off all kinds of culturally-dependent expectations and responses. Thus, owing to her unique ability to support natural communication, we believe Andosan and androids like her constitute a new – but highly familiar – kind of information medium. They can provide a quality of interaction in our daily lives of which ordinary computers – or even humanoid robots – are incapable. Therefore, we see androids one day filling many human roles, such as a guide at a train station or a friend and tutor in an elementary school, which only people have previously been able to perform in a humanlike way.

We would like to emphasize the importance of integrating androids with other new media, such as the Internet, the cellular phone network, the global positioning system, and biometric and identification tagging systems. For example, in our laboratory we have developed a sensor network that integrates data from hundreds of floor sensors, infrared detectors, microphones, and omnidirectional cameras. This information infrastructure provides a robust means of supporting the activities of androids in our office.

It is well known that the early pioneers of AI believed the essence of intelligence to be found in such abstract tasks as proving theorems or playing chess. However, the past fifty years have taught us that the sensorimotor domain is more challenging. Our ancestors spent vastly more time evolving sensorimotor systems that could link with the environment robustly than on developing the so-called higher competencies of language and abstract reason 3) 4). Computer vision approaches that enable a robot to find its bearing and detect individuals and objects still break down outside the laboratory in what is the whirly burly of human existence. Therefore, a sensor rich environment may be key for androids to achieve a high level of performance in practical situations like a crowded shopping mall or cinema. This kind of environment could finesse problems that Mother Nature had more time to solve "on board."

As mentioned earlier, certain lines of scientific inquiry can only be achieved with androids. One example is how appearance and behavior interrelate in forming a subject's perception of human presence[27)] (**Fig. 2**). By employing robots with varying degrees of human likeness and behavior, we can evaluate how these factors influence subjects' experiences and responses – both by questionnaires and other subjective means of evaluation[34)] and through objective measures (such as the distribution of eye fixations as tracked with an Eyemark Recorder). Mori[28)] predicted that people might be frightened of machines that appeared too lifelike or find them unsettling inasmuch as they resemble corpses or zombies. This he refers to as an uncanny valley. However, what we have discovered is that a humanlike appearance elicits the expectation of humanlike behavior. If the android's behavior matches its appearance, the uncanny valley disappears. These effects are also age dependent. A one year old is untroubled by the child replica, even when its behavior is robot-like. Yet a normal three year old will find this combination of appearance and behavior unnerving.

3. A constructivist approach to understanding social mechanisms

Androids can help us do much more than just discover how people relate to different kinds of machines. Owing to their resemblance to humans, android research has the potential to contribute to an understanding of human behavior and the roles our brains and bodies play in supporting it. While neuroscience typically takes a bottom-up approach to studying brains by analyzing them into parts and accumulating knowledge about each part and its interrelations (e.g., the hippocampus or cerebellum), simulating human behavior in androids provides a top-down, synthetic methodology for positing and testing behavior-generating mechanisms at a functionally more abstract level. Approaches that examine theories of neural function by evaluating the behavior they generate when embedded in a humanoid robot 35) can be extended to androids. This is useful because of their unique potential for exploring mechanisms that support humanlike communication among people.

We have proposed a constructivist approach to developing and analyzing cognitive models whereby models are implemented in humanoids, their faults are diagnosed, and then the models are improved and reimplemented 1). This process iterates repeatedly. The Total Turing Test provides the ultimate method of evaluating models because a judge would have to find an android's appearance, behavior and, in some forms of the test, even its internal workings to be indistinguishable from a person's 16) 18) 19) 20) 21). Turing's original imitation game was devised to evaluate the intelligence of computers under the assumption that mental capacities could be abstracted from embodiment 36). This begs many questions about the nature of mental capacities and how internal representations are to be grounded in external states of affairs 17) 23) 24). The Total Turing Test acknowledges that we have good cause to build androids (or what Turing condescendingly refers to as "people suits") because embodiment has proven to be essential to being human. While as a test of intelligence all versions of the imitation game may be flawed 11) 12), a modified Total Turing Test could be used to compare how true-to-life responses constrained by different cognitive models are 7). We have previously referred to this as a *communion game*. It not only provides a means of scrutinizing models within a research project and finding alternatives but of comparing the results from different research teams.

Androids not only provide the ultimate test bed for evaluating cognitive and behavioral theories but a platform for their eventual unification. Since androids require us to confront issues surrounding both mechanism and behavior, we can no longer view cognition as solely a property of brains, to be understood at a micro-structural level, nor as socially-definable and separable from biomechanical or sensorimotor constraints. In other words, androids have the potential for helping researchers to bridge the gap between cognitive neuroscience and the behavioral sciences, leading to a new way of understanding individuals that differs from our current understandings in the social and life sciences. Thus, we hope to find principles underlying the relationship between brains and social activity that will apply equally well to androids and *Homo sapiens*.

A possible framework for these principles has been developing with the common theme of distributed cognition[6)], which has roots in the philosophy of Dennett 8) 9) 31): "Biological systems of the *H. sapiens* variety turn themselves into people – socially embedded teleological selves with narrated biographies in terms of these very beliefs and desires – by taking the intentional stance toward themselves. They can do this thanks to the existence, out in the environment, of public languages that anchor their interpretations to relatively consistent and socially enforced rules of continuity.... [T]hey are incentivized to narrate themselves as coherent and relatively predictable characters, and to care deeply about the dramatic trajectories of these characters they become... [People] are partly constituted out of their social environments, both in the networks of expectations that give identity to them as people, and in the fact that the meanings of their own thoughts are substantially controlled by semantic systems that are collective rather than individual. They are thus not identical to their nervous systems, which are indeed constituted internally." 33)

These ideas illustrate why the development of androids is beyond the scope of mere engineering: To make the android humanlike, we must

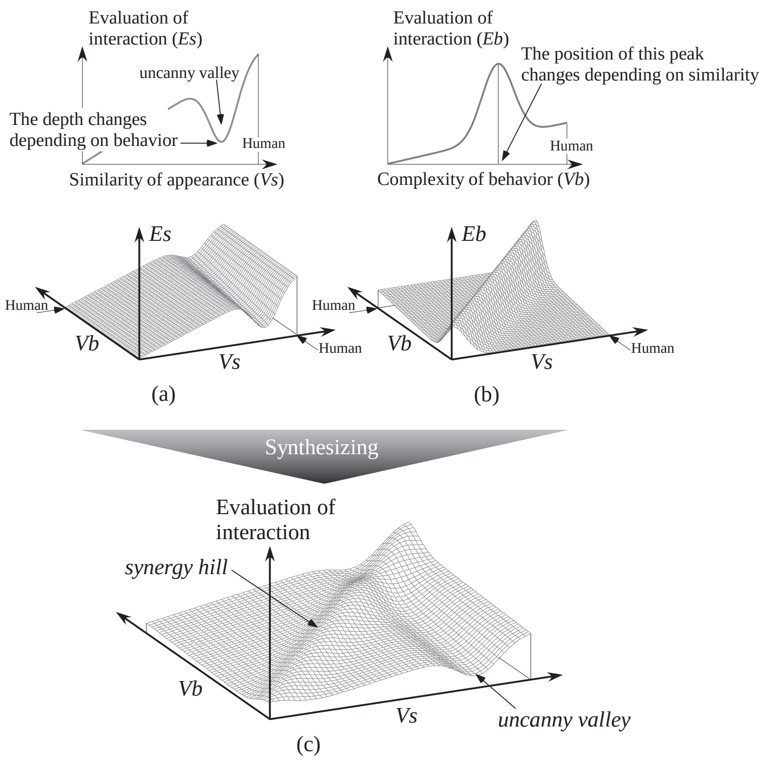

Fig. 2 Mori 28) noticed that, the more humanlike a robot is, the more familiar it seems, until it reaches a level of likeness at which subtle imperfections in appearance and behavior make it repulsive. In a study of gaze behavior, we have attempted to isolate the effects of appearance from those of behavior 27). We hypothesize that *(1)* the uncanny valley in our evaluation of a robot's interaction depends mainly on its appearance but that, *(2)* when appearance and behavior are well-matched in their degree of human-likeness, there is a synergy effect 14). A synthesis of the two hypothesis may imply that a robot's unsettling appearance could be mitigated, if its behavior were very humanlike.

investigate human activity, and to evaluate theories of human activity accurately, we need to implement them in an android. This would appear to call for a new field of inquiry that integrates the constructivist approach from robotics with the empirical methodologies of the social sciences.

4. Potential pitfalls

Roboticists are generally trained in engineering or computer science; however, most of us are amateurs at studying behavior, whether it be human or robotic. I refer not only to the methodological issues of coding and statistical controls, but also broader issues about what to look for, what is known in the literature, and so on. Thus, training in the social sciences, or at least collaboration with social scientists, should be brought to bear on the development of androids.

Unfortunately, social scientists do not typically study social mechanisms that we can implement. For example, a common theory in psychology claims that a cold climate is a mechanism that stimulates eye contact. It is derived from the inverse correlation between mean temperatures and levels of eye contact in various cultures. Even if we set aside the fact that correlation is not causation, obviously this sort of "mechanism" is of little or no use in controlling an android. Social scientists often do not even address questions concerning how a social phenomenon functions. Grahe and Bernieri 15), for example, examine the relative importance of auditory, visual, and text-based information in making accurate judgments about rapport between third parties. Rapport denotes a sense of mutual trust, harmony, sympathy, and friendli-

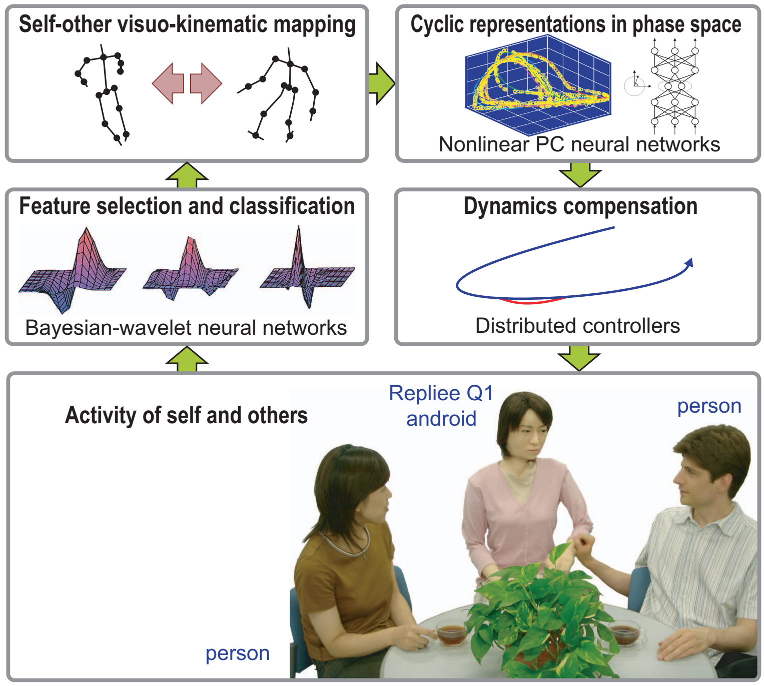

Fig. 3 We would like Andosan, formally referred to as *Repliee Q1*, to learn responses in everyday situations. This involves implementing a mimesis loop. Andosan is seated between two people taking tea. The android recognizes the body parts of others and maps them onto its own body. (This is presently accomplished with the aid of a motion capture system.) Open and closed-curve NLPCA neural networks recognize the motion patters of self and others and generate motion patterns 26) 25). Dynamics are added to compensate for unexpected perturbations. Thus, the android recognizes the behavior of others because it has grounded it in terms of its own body.

ness, which our social robots are meant to engender 22). Although the study demonstrates the relative importance of visual cues, it does not explain how they function as signs of rapport. There are, however, notable exceptions such as Bechinie and Grammer's system for inferring personality from dance 2). His system models, at least implicitly, the relationship between predispositions and perceived behavior.

Mechanisms that attempt to explain how "brains make behavior" are especially risky in the neurosciences because there are far more ways of being provably wrong about brain function than there are of being right.☆ Churchland 5) notes that too few neuroscientists are willing to take these risks, citing Pelionz and Llinias 30) as exceptions for proposing a tensor theory of the cerebellum, which was initially met with strong criticism. Nevertheless, we must posit mechanisms with the understanding that neural verisimilitude is not yet feasible with present day computers. To attempt to simulate on one (or even many) Von Neumann machine the local and parallel computations of whole brains, which are composed of a massive number of highly interconnected computing elements, would lead to a computational bottleneck more severe than the well known Turing tar pit. In addition, a focus on the brain may give a skewed understanding of cognition, which involves social and cultural processes that are active in the relations between individuals. Therefore, roboticists need to use the empirical methodologies of social scientists to help them posit and evaluate social mechanisms that need not be reduced to brain states.

Theories from the social sciences, however, often have other limitations. One difficulty is that they tend to be descriptive – that is, based on descriptions of behavior made in human lan-

☆ In many cases, however, behavior is not "made by brains" but may arise as people respond to each other while exploiting how their activity is orchestrated.

guages – rather than being explanatory in the sense of lending themselves to android implementation. The historical mistake of artificial intelligence is to confound these descriptions, and especially descriptions from our folk psychology, with the mechanisms that enable persons to orient toward them – and to try to encode these descriptions in symbol systems that encode their posited roles in making inferences 29) 17) 23).

Similar difficulties occur in procedural (as opposed to deductive) symbol systems, which involve planning and selecting actions 10). Although we may conceive of human activity in terms of how culturally-mediated forms (e.g., language) have biased our categorical perceptions, to reify categories of perceived action and abstract them from sensorimotor relations is suspect 24). If we could subtract easily labeled or consciously recognizable actions from behavior, this would leave a tremendous residue of unlabeled or only subconsciously recognized activity. The idea that we can implement the fullness of human activity by decomposing it into *a priori* action-categories may be mistaken. Arguably, there are no actions, only action. Action is the result of interactions among dynamic systems and occurs because of a closed circuit of recognition and response at many spatial and temporal scales. The brain implements numerous kinds of constraints on action, but the actions we imagine ourselves to be performing are in some sense illusory – the outcome our own categorization and language processes or, to put it another way, taking an intentional stance to ourselves.

Fortunately, the embodiment of social and cognitive theories in interactive robots sets a much higher bar than text-based AI systems for theory evaluation. Theories that reify descriptions that rely on a human interpreter for their grounding cannot be implemented in autonomous systems 23). The demands of coherently integrating responses cross-modally and coping with open, socially complex environments limit the applicability of descriptive theories. Androids will be confronted with circumstances that exhibit complex closely-coordinated social dynamics, where stable patterns emerge at various spatial and temporal scales, and expectations depend in part on a histories of interaction that are unique to individual relationships. It is out of these social circumstances that androids must construct themselves as social beings just as human beings have constructed themselves into people.

5. Conclusion

In sum, since the development of androids and the investigation of the mechanisms underlying social activity are mutually dependent, we need to found a new, cross-disciplinary field that we have dubbed *android science*. Only a humanlike appearance and behavior can elicit humanlike communication. Our research should maintain its unique character and focus while integrating insights from cognitive and brain science, the social and behavioral sciences, robotics, sensor information processing, material science, mechanical and control engineering, and artificial intelligence. We can never deeply understand human beings without building androids, and we cannot build androids without deepening our understanding of what it means to be human.

Acknowledgments We have developed the android robots *Repliee R1* and *Repliee Q1* (which we informally call Lisa and Andosan) in collaboration with Kokoro Company, Ltd.

Fundamental research is supported by several grants from the Japanese Government: the Advanced and Innovational Research Program in Life Sciences and Grant-in-Aid for Scientific Research on Priority Areas "Understanding Human Information Processing and Its Applications" from the Ministry of Education, Culture, Sports, Science, and Technology and the Next Generation Robot Project grant from the New Energy and Industrial Technology Development Organization.

References

1) M. Asada, K.F. MacDorman, H. Ishiguro, and Y. Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous System*, 37:185–193, 2001.
2) M. Bechinie and K. Grammer. Charisma Cam: A prototype of an intelligent digital sensory organ for virtual humans. In *Intelligent Virtual Agents*, Lecture Notes in Artificial Intelligence 2792, pages 212–216, Berlin, 2003. Springer-Verlag.
3) R.A. Brooks. Intelligence without reason. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, volume 1, pages 569–595. Morgan Kaufmann, 1991.
4) R.A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

5) P.S. Churchland. *Neurophilosophy: Toward a Unified Science of the Mind-Brain.* MIT Press, Cambridge, MA, 1986.
6) A. Clark. *Being There: Putting Brain, Body and World Together Again.* MIT Press, Cambridge, MA, 1997.
7) S.J. Cowley and K.F. MacDorman. Simulating conversations: The communion game. *AI & Society*, 9:116–137, 1995.
8) D.C. Dennett. *The Intentional Stance.* MIT Press, Cambridge, MA, 1987.
9) D.C. Dennett. *Consciousness Explained.* Little Brown, Boston, MA, 1991.
10) R. Fikes and N. Nilsson. Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208, 1971.
11) R.M. French. Subcognition and the limits of the Turing Test. *Mind*, 99:53–65, 1990.
12) R.M. French. The Turing Test: The first fifty years. *Trends in Cognitive Sciences*, 4:115–121, 2000.
13) J.J. Gibson. *The ecological approach to visual perception.* Houghton Mifflin, Boston, MA, 1979.
14) J. Goetz, S. Kiesler, and A. Powers. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The Twelveth IEEE International Workshop on Robot and Human Interactive Communication*, Lisbon, Portugal, 31 October–2 November 2003.
15) J.E. Grahe and F.J. Bernieri. The importance of nonverbal cues in judging rapport. *Journal of Nonverbal Behavior*, 23:253–269, 1999.
16) S. Harnad. Minds, machines and Searle. *Journal of Experimental and Theoretical Artificial Intelligence*, 1:5–25, 1989.
17) S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.
18) S. Harnad. Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, 1:43–54, 1991.
19) S.Harnad. The Turing Test is not a trick: Turing indistinguishability is a scientific criterion. *SIGART Bulletin*, 3:9–10, 1992.
20) S. Harnad. Levels of functional equivalence in reverse bioengineering: The Darwinian Turing Test for artificial life. *Artificial Life*, 1:293–301, 1994.
21) S. Harnad. Minds, machines and Turing: The indistinguishability of indistinguishables. *Journal of Logic, Language, and Information*, 9:425–445, 2000.
22) T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1-2):61–84, 2004.
23) K.F. MacDorman. *Symbol grounding: Learning categorical and sensorimotor predictions for coordination in autonomous robots (Technical Report No. 423).* Computer Laboratory (e-mail librarian@cl.cam.ac.uk for a copy), Cambridge, 1997.
24) K.F. MacDorman. Grounding symbols through sensorimotor integration. *Journal of the Robotics Society of Japan*, 17:20–24, 1999.
25) K.F. MacDorman, R. Chalodhorn, and M. Asada. Periodic nonlinear principal component neural networks for humanoid motion segmentation, generalization, and generation. In *ICPR-2004: International Conference on Pattern Recognition*, Cambridge, UK, August 23–26 2004.
26) K.F. MacDorman, R. Chalodhorn, and H. Ishiguro. Learning to recognize and reproduce abstract actions from proprioception. In *The Third International Conference on Development and Learning: Developing Social Brains*, Salk Institute, La Jolla, California, 20–22 October 2004.
27) T. Minato, M. Shimada, H. Ishiguro, and S. Itakura. Development of an android robot for studying human-robot interaction. In R.Orchard, C. Yang, and M. Ali, editors, *Innovations in Applied Artificial Intelligence*, pages 424–434, Berlin, 2004.
28) M. Mori. *The buddha in the robot: A robot engineer's thought on science and religion.* Tuttle Publishing, Tokyo, 1982.
29) A. Newell. Physical symbol systems. *Cognitive Science*, 4:135–183, 1980.
30) A. Pellionisz and R.R. Llinas. Tensor network theory of the metaorganization of functional geometries in the central nervous system. *Neuroscience*, 16:245–273, 1985.
31) P. Pettit. *The Common Mind.* Oxford University Press, Oxford, 1993.
32) B. Reeves and C. Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press (CSLI), New York, 1996.
33) D. Ross. Emotions as strategic signals. *Rationality and Society*, 16:251–286, 2004.
34) G. Rugg and P. McGeorge. *Questioning Methodology.* University College Northampton Working Paper, Northamption, England, 1999.
35) N. Schweighofer, M. Arbib, and M. Kawato. Role of the cerebellum in reaching movements in humans. I. Distributed inverse dynamics control. *European Journal of Neuroscience*, 10:86–94, 1998.
36) A.M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.