AISB/IACAP World Congress 2012

Birmingham, UK, 2-6 July 2012

REVISITING TURING AND HIS TEST: COMPREHENSIVENESS, QUALIA, AND THE REAL WORLD

Vincent C. Müller and Aladdin Ayesh (Editors)

Part of

The Chairs
Vincent C. Müller (Anatolia College/ACT & University of Oxford) and
Aladdin Ayesh (De Montfort University)

With the Support of:
Mark Bishop (Goldsmiths, University of London),
John Barnden (University of Birmingham),
Alessio Plebe (University Messina) and
Pietro Perconti (University Messina)

The Program Committee:
Raul Arrabales (Carlos III University of Madrid),
Antonio Chella (University of Palermo),
Giuseppe Trautteur (University of Napoli Federico II),
Rafal Rzepka (Hokkaido University)
… plus the Organizers Listed Above

The website of our symposium is on http://www.pt-ai.org/turing-test

Interactive Intelligence: Behaviour-based AI, Musical HCI and the Turing Test

Adam Linson, Chris Dobbyn and Robin Laney

Abstract. The field of behaviour-based artificial intelligence (AI), with its roots in the robotics research of Rodney Brooks, is not predominantly tied to linguistic interaction in the sense of the classic Turing test (or, "imitation game"). Yet, it is worth noting, both are centred on a behavioural model of intelligence. Similarly, there is no intrinsic connection between musical AI and the language-based Turing test, though there have been many attempts to forge connections between them. Nonetheless, there are aspects of musical AI and the Turing test that can be considered in the context of non-language-based interactive environments–in particular, when dealing with *real-time* musical AI, especially interactive improvisation software. This paper draws out the threads of *intentional agency* and *human indistinguishability* from Turing's original 1950 characterisation of AI. On the basis of this distinction, it considers different approaches to musical AI. In doing so, it highlights possibilities for non-hierarchical interplay between human and computer agents.

1 Introduction

The field of behaviour-based artificial intelligence (AI), with its roots in the robotics research of Rodney Brooks, is not predominantly tied to linguistic interaction in the sense of the classic Turing test (or, "imitation game" [24]). Yet, it is worth noting, both are centred on a behavioural model of intelligence. Similarly, there is no intrinsic connection between musical AI and the language-based Turing test, though there have been many attempts to forge connections between them. The primary approach to applying the Turing test to music is in the guise of so-called "discrimination tests", in which human- and computer-generated musical output are compared (for an extensive critical overview of how the Turing test has been applied to music, see [1]). Nonetheless, there are aspects of musical AI and the Turing test that can be considered in the context of non-language-based interactive environments—in particular, when dealing with *real-time* musical AI, especially interactive improvisation software (see, for example, [23] and [8]). In this context, AI for non-hierarchical human-computer musical improvisation such as George Lewis' *Voyager* [16] and Turing's imitation game are both examples of "an open-ended and performative interplay between [human and computer] agents that are not capable of dominating each other" [21].

2 Background

It is useful here to give some context to the Turing test itself. In its original incarnation, the test was proposed as a thought experiment to explain the concept of a thinking machine to a public uninitiated

Faculty of Mathematics, Computing and Technology, Dept. of Computing, Open University, UK. Email: a.linson, c.h.dobbyn, r.c.laney @open.ac.uk

in such matters [24]. Rather than as a litmus test of whether or not a machine could think (which is how the test is frequently understood), the test was in fact designed to help make sense of the concept of a machine that could think. Writing in 1950, he estimates "about fifty years' time" until the technology would be sufficient to pass a real version of the test and states his belief "that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted". Thus his original proposal remained a theoretical formulation: in principle, a machine could be invented with the capacity to be mistaken for a human; if this goal were accomplished, a reasonable person should accept the machine as a thinking entity. He is very clear about the behaviourist underpinnings of the experiment:

> May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.

He goes on to describe the "imitation game" as one in which the machine should "try to provide answers that would naturally be given by a man". His ideas became the basis for what eventually emerged as the field of AI.

As Turing emphasised, the thought experiment consisted of an abstract, "imaginable" machine that—under certain conditions to ensure a level playing field—would be indistinguishable from a human, from the perspective of a human interrogator [24]. Presently, when the test is actually deployed in practice, it is easy to forget the essential role of the designer, especially given the fact that the computer "playing" the game is, to an extent, thrust into the spotlight. In a manner of speaking, the interactive computer takes centre stage, and attention is diverted from the underlying challenge set forth by Turing: *to determine the specifications* of the machine. Thus, one could say in addition to being a test for a given machine, it is also a creative design challenge to those responsible for the machine. The stress is on design rather than implementation, as Turing explicitly suggests imagining that any proposed machine functions perfectly according to its specifications (see [24], p. 449). If the creative design challenge were fulfilled, the computer would behave convincingly as a human, perhaps hesitating when appropriate and occasionally refusing to answer or giving incorrect answers such as the ones Turing imagines [24]:

> Q: Please write me a sonnet on the subject of the Forth Bridge.
> A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.
A: (Pause about 30 seconds and then give as answer) 105621.

The implication of Turing's example is that the measure of success for those behind the machine lies in designing a system that is also as stubborn and fallible as humans, rather than servile and (theoretically) infallible, like an adding machine.

3 Two threads unraveled

Two threads can be drawn out of Turing's behavioural account of intelligence that directly pertain to contemporary AI systems: the first one concerns the kind of intentional agency suggested by his example answer, "count me out on this one"; the second one concerns the particular capacities and limitations of human embodiment, such as the human inability to perform certain calculations in a fraction of a second and the human potential for error. More generally, the second thread has to do with the broadly construed linguistic, social, mental and physical consequences of human physiology. Indeed, current theories of mind from a variety of disciplines provide a means for considering these threads separately. In particular, relevant investigations that address these two threads—described in this context as *intentional agency* and *human indistinguishability*—can be found in psychology, philosophy and cognitive science.

3.1 Intentional agency

The first thread concerns the notion of intentional agency, considered here separately from the thread of human indistinguishability. Empirical developmental psychology suggests that the human predisposition to attribute intentional agency to both humans *and* nonhumans appears to be present from infancy. Poulin-Dubois and Shultz chart childhood developmental stages over the first three years of life, from the initial ability to identify agency (distinguishing animate from inanimate objects) on to the informed attribution of intentionality, by inference of goal-directed behavior [22]. Csibra found that infants ascribed goal-directed behavior even to artificially animated inanimate objects, if the objects were secretly manipulated to display teleological actions such as obstacle avoidance [7]. Király, et al. identify the source of an infant's interpretation of a teleological action: "if the abstract cues of goal-directedness are present, even very young infants are able to attribute goals to the actions of a wide range of entities even if these are unfamiliar objects lacking human features" [10].

It is important to note that in the above studies, the infants were passive, remote observers, whereas the Turing test evaluates direct interaction. While the predisposition of infants suggests an important basis for such evaluation, more is needed to address interactivity. In another area of empirical psychology, a study of adults by Barrett and Johnson suggests that even a lack of apparent goals by a self-propelled (nonhuman) object can lead to the attribution of intentionality in an interactive context [2]. In particular, their test subjects used language normally reserved for humans and animals to describe the behaviour of artificially animated inanimate objects that appeared to exhibit resistance to direct control in the course of an interaction; when there was no resistance, they did not use such language. The authors of the study link the results of their controlled experiment to the anecdotal experience of the frustration that arises during interactions with artifacts such as computers or vehicles that "refuse" to cooperate. In other words, in an interactive context, too much passivity by an artificial agent may negate any sense of its apparent

intentionality. This suggests that for an agent to remain apparently intentional during direct interaction, it must exhibit a degree of resistance along with the kind of adaptation to the environment that indicates its behaviour is being adjusted to attain a goal. These features appear to be accounted for in Turing's first example answer above: the answer is accommodating insofar as it is a direct response to the interrogator, but the show of resistance seems to enhance the sense of "intelligence". It is noteworthy that this particular thread, intentional agency, relates closely to Brooks' extension of intelligence to nonlinguistic, nonhuman intelligence, especially in relation to insect and other animal intelligence, which he has emulated in robotic form with his particular approach to AI (see [3]).

3.2 Human indistinguishability

The second thread, the idea that human capacities and limitations should be built into an AI system, strongly relates to many significant accounts of embodied, situated activity (see, for example, [9], [4] and [11]). These accounts focus on how the human body, brain, mind and environment fundamentally structure the process of cognition, which can be understood through observable behaviour. When dealing with AI, the focus on behaviour clearly ties back to Turing. These themes are also taken up in Brooks' behaviour-based AI approach, but, at least in his early research, he applies them primarily to nonhuman intelligence. In particular, he relates these themes to the kinds of adaptive behaviour described in the first thread. The differing properties of the second thread will come into sharper focus by returning to Turing's example, for a consideration of matters particular to humans.

Although Turing's example of pausing and giving an incorrect answer is a clear example of a human limitation over a machine, it is possible to give an inverted example of human and machine competence that applies equally well. If the question posed to the machine were instead "Is it easy to walk from here to the nearest supermarket?", the machine's answer would depend on how its designers handled the notion of "easy to walk to". In this case, the machine must not only emulate humans' abstract cognitive limitations when solving arithmetical problems; it must also be able to respond according to human bodily limitations. One could easily imagine a failed machine calculation: the supermarket is at the end of a single straight road, with no turns; it answers "yes, it is easy to walk to". But if the supermarket is very distant, or nearby but up a steep incline, then in order for the machine to give an answer that is indistinguishable from a human one, it must respond in a way that seems to share our embodied human limitations. Returning to the arithmetic example, as Doug Lenat points out, even some wrong answers are more human than others: "   is more understandable than if the program pretends to get a wrong answer of 0 or   for that subtraction problem" [14]. Although Lenat disputes the need for embodiment in AI (he prefers a central database of human common sense [13], which could likely address the "easy to walk to" example), it could be argued, following the above theoretical positions, that the set of humanlike wrong answers is ultimately determined by the "commonalities of our bodies and our bodily and social experience in the world" [11].

This second thread, which could also be characterised as *the attempt to seem humanlike*, is taken up in another nonlinguistic area of AI, namely, musical AI. Some "intelligent" computer music composition and performance systems appear very close to achieving human indistinguishability in some respects, although this is not always their explicitly stated purpose. For example, Manfred Clynes

describes a computer program that performs compositions by applying a single performer's manner of interpretation to previously unencountered material, across all instrumental voices [5]. He states that "our computer program plays music so that it is impossible to believe that no human performer is involved," which he qualifies by explaining the role of the human performer as a user of the software, who "instills the [musical performance] principles in the appropriate way". Taking an entirely different approach, David Cope, argues that a Turing-like test for creativity would be more appropriate to his work than a Turing test for intelligence [6]. On the other hand, he has called his well-known project "Experiments in Musical Intelligence" and he also makes reference to "intelligent music composition". Furthermore, he states that his system generates "convincing" music in the style of a given composer (by training the system with a corpus of human-composed music), and one can infer that, in this context, "convincing" at least approximates the notion of human indistinguishability. With a more critical articulation, Pearce and Wiggins carefully differentiate between a test for what Cope calls "convincing" and a Turing test for intelligence [19]. As they point out, despite the resemblance of the two approaches, testing for intelligence is distinct from determining the "(non-)membership of a machine composition in a set of human composed pieces of music". They also note the significant difference between an interactive test and one involving passive observation.

4 Broadening the interactive horizon

One reason for isolating these two threads is to recast Turing's ideas in a wider social context, one that is better attuned to the contemporary social understanding of the role of technology research: namely, that it is primarily intended (or even expected) to enhance our lives. Outside the thought experiment, in the realm of practical application, one might redirect the resources for developing a successful Turing test candidate (e.g., for the Loebner Prize) and instead apply them toward a different kind of interactive system. This proposed system could be built so that it might be easily identified as a machine (even if occasionally mistaken for a human), which seemingly runs counter to the spirit of the Turing test. However, with an altered emphasis, one could imagine the primary function of such a machine as *engaging* humans in a continuous process of interaction, for a variety of purposes, including (but not limited to) stimulating human creativity and providing a realm for aesthetic exploration.

One example of this kind of system is musical improvisation software that interacts with human performers in real time, in a mutually influential relationship between human and computer, such as Lewis' *Voyager*. In his software design, the interaction model strongly resembles the way in which Turing describes a computer's behaviour: it is responsive, yet it does not always give the expected answer, and it might interrupt the human interlocutor or steer the interaction in a different direction (see [16]). In the case of an interactive improvising music system, the environment in which the human and computer interact is not verbal conversation, but rather, a culturally specific aesthetic context for collaborative music-making. In this sense, a musical improvisation is not an interrogation in the manner presented by Turing, yet "test" conversations and musical improvisations are examples of free-ranging and open-ended human-computer interaction. Among other things, this kind of interaction can serve as a basis for philosophical enquiry and cognitive theory that is indeed very much in the spirit of Turing's 1950 paper [24] (see also [15] and [17]).

Adam Linson's *Odessa* is another intelligent musical system that is similarly rooted in freely improvised music (for a detailed descrip-

tion, see [18]). It borrows from Brooks' design approach in modelling the behaviour of an intentional agent, thus clearly taking up the first thread that has been drawn out here. Significantly, it isolates this thread (intentional agency) for study by abstaining from a direct implementation of many of the available methods for human emulation (aimed at the second thread), thus resulting in transparently nonhuman musical behaviour. Nonetheless, initial empirical studies suggest that the system affords an engaging and stimulating human-computer musical interaction. As the system architecture (based on Brooks' subsumption architecture) is highly extensible, future iterations of the system may add techniques for approximating fine-grained qualities of human musicianship. In the meantime, however, further studies are planned with the existing prototype, with the aim of providing insights into aspects of human cognition as well as intelligent musical agent design.

5 Conclusion

Ultimately, whether an interactive computer system is dealing with an interrogator in the imitation game or musically improvising with a human, the system must be designed to "respond in lived real time to unexpected, real-world input" [17]. This responsiveness takes the form of what sociologist Andrew Pickering calls the "dance of agency", in which a reciprocal interplay of resistance and accommodation produces unpredictable emergent results over time [20]. This description of a sustained, continuous play of forces that "interactively stablize" each other could be applied to freely improvised music, whether performed by humans exclusively, or by humans and computers together. Pickering points out a concept similar to the process of interactive stabilisation, 'heterogeneous engineering', elaborated in the work of his colleague John Law (see [12]); the latter, in its emphasis on productive output, is perhaps more appropriate to the musical context of free improvisation.

Although these theoretical characterisations may seem abstract, they concretely pertain to the present topic in that they seek to address the "open-ended and performative interplay between agents that are not capable of dominating each other" [21], where the agents may include various combinations of humans, computers and other entities, and the interplay may include linguistic, musical, physical and other forms of interaction. With particular relevance to the present context, Pickering applies his conceptual framework of agent interplay to the animal-like robots of Turing's contemporary, cybernetics pioneer Grey Walter, and those of Brooks, designed and built decades later [21]. Returning to the main theme, following Brooks, "the dynamics of the interaction of the robot and its environment are primary determinants of the structure of its intelligence" [3]. Thus, independent of its human resemblance, an agent's ability to negotiate with an unstructured and highly dynamic musical, social or physical environment can be treated as a measure of intelligence closely aligned with what Turing thought to be discoverable with his proposed test.

REFERENCES

[1] C. Ariza, 'The interrogator as critic: The turing test and the evaluation of generative music systems', *Computer Music Journal*, 33(2), 48–70, (2009).

[2] J.L. Barrett and A.H. Johnson, 'The role of control in attributing intentional agency to inanimate objects', *Journal of Cognition and Culture*, 3(3), 208–217, (2003).

[3] R.A. Brooks, *Cambrian intelligence: the early history of the new AI*, MIT Press, 1999.

[4] A. Clark, *Being There: Putting Brain, Body, and World Together Again*, MIT Press, 1997.

[5] M. Clynes, 'Generative principles of musical thought: Integration of microstructure with structure', *Communication and Cognition AI, Journal for the Integrated Study of Artificial Intelligence, Cognitive Science and Applied Epistemology*, 3(3), 185–223, (1986).

[6] D. Cope, *Computer Models of Musical Creativity*, MIT Press, 2005.

[7] G. Csibra, 'Goal attribution to inanimate agents by 6.5-month-old infants', *Cognition*, 107(2), 705–717, (2008).

[8] R.T. Dean, *Hyperimprovisation: Computer-interactive sound improvisation*, AR Editions, Inc., 2003.

[9] H. Hendriks-Jansen, *Catching ourselves in the act: Situated activity, interactive emergence, evolution, and human thought*, MIT Press, 1996.

[10] I. Király, B. Jovanovic, W. Prinz, G. Aschersleben, and G. Gergely, 'The early origins of goal attribution in infancy', *Consciousness and Cognition*, 12(4), 752–769, (2003).

[11] G. Lakoff and M. Johnson, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*, Basic Books, 1999.

[12] J. Law, 'On the social explanation of technical change: The case of the portuguese maritime expansion', *Technology and Culture*, 28(2), 227–252, (1987).

[13] D.B. Lenat, 'Cyc: A large-scale investment in knowledge infrastructure', *Communications of the ACM*, 38(11), 33–38, (1995).

[14] D.B. Lenat, 'The voice of the turtle: Whatever happened to ai?', *AI Magazine*, 29(2), 11, (2008).

[15] G. Lewis, 'Interacting with latter-day musical automata', *Contemporary Music Review*, 18(3), 99–112, (1999).

[16] G. Lewis, 'Too many notes: Computers, complexity and culture in voyager', *Leonardo Music Journal*, 33–39, (2000).

[17] G. Lewis, 'Improvising tomorrow's bodies: The politics of transduction', *E-misférica*, 4.2, (2007).

[18] A. Linson, C. Dobbyn, and R. Laney, 'Improvisation without representation: artificial intelligence and music', in *Proceedings of Music, Mind, and Invention: Creativity at the Intersection of Music and Computation*, (2012).

[19] M. Pearce and G. Wiggins, 'Towards a framework for the evaluation of machine compositions', in *Proceedings of the AISB*, pp. 22–32, (2001).

[20] A. Pickering, *The mangle of practice: Time, agency, and science*, University of Chicago Press, 1995.

[21] A. Pickering, *The cybernetic brain: Sketches of another future*, University of Chicago Press, 2010.

[22] D. Poulin-Dubois and T.R. Shultz, 'The development of the understanding of human behavior: From agency to intentionality', in *Developing Theories of Mind*, eds., Janet W. Astington, Paul L. Harris, and David R. Olson, 109–125, Cambridge University Press, (1988).

[23] R. Rowe, *Machine musicianship*, MIT Press, 2001.

[24] A.M. Turing, 'Computing machinery and intelligence', *Mind*, 59(236), 433–460, (1950).