This is the accepted draft of: Proudfoot D. (2014) *Turing's Three Senses of 'Emotional'.* International Journal of Synthetic Emotions 5(2): 7-20. http://dx.doi.org/10.4018/ijse.2014070102. In citing, please cite the published version.

Turing's Three Senses of "Emotional"

Diane Proudfoot

Department of Philosophy, University of Canterbury, New Zealand

Abstract

Turing used the expression "emotional" in three distinct ways: to state his philosophical theory of the concept of intelligence, to classify arguments for and against the possibility of machine intelligence, and to describe the education of a "child machine". His remarks on emotion include several of his most important philosophical claims. This paper analyses these remarks and their significance for current research in Artificial Intelligence.

Keywords

Affective Computing, Anthropomorphism, Child Machine, Emotional Concept, P-Type, Response-Dependence, Turing, Turing Test

Turing's Three Senses of "Emotional"

Introduction[1]

Philosophical and psychological theories of emotion attempt to explain the nature and origin of emotion; in particular, they focus on the rationality of emotion and the role of emotion in deliberative decision-making. This is complicated by the fact that in everyday language the word "emotion" has multiple uses. These different senses have prompted conflicting analyses of emotion—as a feeling (quale or somatic state), as behaviour, or as a cognitive state. Consistent with the multiplicity of everyday uses of "emotion" and its cognates, Turing used the expression "emotional" in three quite distinct ways—to discuss respectively emotional concepts, emotional arguments, and emotional communication. In the first case he used this expression to mean (what philosophers call) *response-dependent*,[2] in the second case he used it to mean *irrational*, and in the third case he used it (tongue-in-cheek) to mean *feeling*.

Turing's notions of emotional concepts, arguments, and communication are central to his philosophy of machine intelligence. *What is intelligence?* Here Turing suggested a novel and intriguing approach to the concept of intelligence that is yet to be developed. *Can machines think?* Turing dissected and countered objections to the possibility of machine intelligence—objections that are still found today. *How are we to build an intelligent machine?* Here Turing set out a research programme for AI—to build a "child machine"—that is now pursued in social and developmental robotics. In this paper I analyse Turing's comments on emotion and their significance for affective computing in the 21st century.

The First Sense—Emotional vs Mathematical

What is intelligence? Turing said that "the idea of intelligence is itself 'emotional' rather than mathematical" and that the concept of intelligence is an "emotional concept" (Turing, 1948, pp. 411, 431). 'Emotional concept' is a term of art for Turing; an emotional concept is a concept the application of which is determined in part by an observer's reactions. He said, "The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration" (Turing, 1948, p. 431).

In modern philosophical terminology such concepts are said to be "response-dependent". Examples are phenomenal concepts such as colour and value concepts such as (moral) goodness or beauty. It would be a mistake, response-dependence theorists argue, to analyse the concept of colour solely in terms of electro-magnetic radiation or the concept of beauty solely in terms of the micro-physical properties of a beautiful object. This is because an object is red or beautiful only if it *looks* red or beautiful. In a response-dependence theory of intelligence of the sort suggested by Turing's remarks, whether or not an entity is intelligent (or thinks) is similarly determined in part by our responses to the entity—the entity must *appear* intelligent.[3] The concept of intelligence (or thinking) is not to be analysed solely in terms of brain processes, at either a physical or computational level. Turing made this point when he said that if "one can see the cause and effect working themselves out in the brain, one regards it as not being thinking, but a sort of unimaginative donkey-work" (Turing, Braithwaite, Jefferson, & Newman, 1952, p. 500).

The thesis that intelligence is a response-dependent concept is the philosophical thesis that underlies Turing's use of his computer-imitates-human game as a "criterion of 'thinking'" in machines (Turing, 1950, p. 436).[4] In each of Turing's three versions of his test, it is the observer's *reaction* to the machine—rather than merely the machine's behaviour—that is crucial. (On the three versions of the test, see Copeland, 2000, 2004; Proudfoot, 2005, 2013a.) Turing described his first version of the imitation game, the 1948 chess-playing game, as a "little experiment" to see if a human observer would "imagine intelligence" in the machine (Turing, 1948, p. 431). The goal of his famous 1950 imitation game is also that the observer (i.e. the interrogator) be fooled by the machine. In fact, Turing's central question is whether the interrogator in this computer-imitates-human game will be fooled as often as the interrogator in a man-imitates-woman game.[5] He said, "Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? [This question replaces] our original, 'Can machines think?'" (Turing, 1950, p. 441).[6] In Turing's 1952 version of the imitation game, the goal is again that the interrogator be "taken in by the pretence" (Turing, Braithwaite, Jefferson, & Newman, 1952, p. 495).

If intelligence is an emotional concept, can we say that intelligence is in the eye of the beholder? These words suggest that intelligence is a subjective concept; on this view, anything goes when it comes to ascribing intelligence to entities.[7] Some modern AI researchers do say that "intelligence is in the eye of the observer"—these are Rodney Brooks' words, for example (Brooks, 1995, p. 57). However, response-dependence theorists claim that the notion of response-dependence leaves room to say that response-dependent concepts are objective, giving us information about the world.[8] Not just *any* observer is qualified to make judgements of response-dependent concepts; an act is morally good or an object beautiful, theorists say, if it appears so to *normal* subjects in *normal* conditions. The same constraint will apply in the case of a response-dependence theory of intelligence (see Proudfoot, 2013a, for Turing's specification of normal subjects and conditions). In addition, Turing said that judging something as intelligent is also determined "by the properties of the object under consideration"—what makes an entity intelligent is not solely the observer's reaction, but is in part the (response-*in*dependent) qualities of the entity itself (Turing, 1948, p. 431). This suggests that for Turing, even though the concept of intelligence is "emotional", intelligence is real.

Turing's distinction between an "emotional" and a "mathematical" concept sets him apart from much current theorizing in AI. Many researchers assume that AI's underlying goal is a computational model of cognition, in particular of intelligence—saying, for example, "AI is the engineering of cognition based on the computational vision which runs through and informs all of cognitive science" (Hayes & Ford, 1995, p. 976). Some researchers point to Turing himself as the source of this idea—saying, for example, that it is Turing who "sends us in search of the … computational procedures that the brain presumably employs in generating its behavioral magic" (Churchland, 2008, p. 109). But this goal is at odds with Turing's own words. The concept of a number (or function) is not an emotional concept: which mathematical (or computational) property a machine has is independent of our reaction to the machine.[9] If the concept of intelligence is an emotional concept, no response-independent property—including computation—can suffice for intelligence or thinking.

The Second Sense—Emotional vs Rational

In the late 1940s and early 1950s in the UK, as people learned about the Manchester "Baby" and the EDSAC, they asked whether machines like these could actually think. Turing was scathing about popular views on the question *Can machines think?*. He said that "the naive point of view of the man in the street"—who might simply "disbelieve what he has heard" of the activities of machines—is "almost superstitious" (Turing, 1951, p. 482). According to Turing, "[m]any people are extremely opposed to the idea of [a] machine that thinks" and this is "simply because they do not like the idea": "If a machine can think, it might think more intelligently than we do, and then where should we be? … [W]e should, as a species, feel greatly humbled" (Turing, 1951, pp. 485-6). The prospect of superhuman-level AI, Turing said, is "remote but not astronomically remote and is certainly something which can give us anxiety" (Turing, 1951, p. 486).

Assessing attitudes to the possibility of machine intelligence, Turing distinguished between (what he described as) "rational" and "irrational" objections (Turing, 1951, p. 485). In his view, some objections to AI are "purely emotional" (Turing, 1948, p. 411). These include the "religious belief that any attempt to construct such machines is a sort of Promethean irreverence" and also the "unwillingness to admit the possibility that mankind can have any rivals in intellectual power" (Turing, 1948, p. 410). The former objection is the "The Theological Objection" (namely, thinking is "a function of man's immortal soul") and the latter is "The 'Heads in the Sand' Objection" (Turing, 1950, pp. 449-50). Purely emotional objections, Turing said, "do not really need to be refuted. If one feels it necessary to refute them there is little to be said that could hope to prevail" (Turing, 1948, p. 411). In this sense of "emotional", an emotional objection is not susceptible to reasoned argument.[10] Turing said:

> We like to believe that Man is in some subtle way superior to the rest of creation. It is best if he can be shown to be *necessarily* superior, for then there is no danger of him losing his commanding position. The popularity of the theological argument is clearly connected with this feeling. … I do not think that this argument is sufficiently substantial to require refutation. Consolation would be more appropriate; perhaps this should be sought in the transmigration of souls. (Turing, 1950, p. 450)

The only reply to an emotional objection is itself emotional—that is, irrational.

According to Turing, even those who allow the possibility of intelligent machines would resist *building* such machines—"unless we have advanced greatly in religious toleration from the days of Galileo" (Turing, c. 1951, p. 475). Even mathematicians working on the ACE "may be unwilling to let their jobs be stolen from them in this way. … [T]hey would surround the whole of their work with mystery and make excuses, couched in well chosen gibberish, whenever any dangerous suggestions were made" (Turing, 1947, p. 392). Purely emotional objections seem for Turing to be nothing more than well-chosen gibberish.

Some current AI researchers regard all manner of philosophical objections to human-level AI as emotional in this sense, stemming from unwarranted sentiment. For example, Ray Kurzweil says of John Searle (and the Chinese room argument) that he has a "biology-centric view of consciousness", a "bias that computers are inherently incapable of 'mental life'", and "a basic lack of understanding of technology" (Kurzweil, 2002, pp. 131, 164, 170). In a similar vein, Robin Hanson complains of Roger Penrose that his famous *Emperors of the Mind* is "a sloppier collection of arguments for what is clearly a deeply-held opinion" (Hanson, 1991). The philosopher John Lucas has complained of this attitude, saying, "Although I argued with what I hope was becoming modesty and a certain degree of tentativeness, many of the replies have been lacking in either courtesy or caution. I must have touched a raw nerve" (Lucas, 1990).

Lucas's argument (and Penrose's) is that Gödel's theorem "shows that minds cannot be explained by any computational system" (Lucas, 2013; see Lucas, 1961, for the original exposition of his argument).[11] Turing himself carefully considered the same argument (based on his own as well as Gödel's results) in his 1948 and 1950 papers. He called it "The Mathematical Objection" and did not consider it irrational (Turing, 1950, pp. 450-1; see also 1948, pp. 410-11). But this does not free Lucas entirely from criticism. He replies as follows to Turing's attack on the Heads in the Sand Objection:

> [T]here are good reasons for being chary of throwing over established modes of thought too easily. They may have much going for them, and often have been tried over many generations, and found to be reliable. (Lucas, 2008, p. 68)

This stance lays Lucas (but not his original objection) open to the charge that he is excessively conservative and lacks imagination.

The Third Sense—Emotional vs Unemotional

How are we to build an intelligent machine? Turing recommended that we begin with a "comparatively simple" machine and give it a "suitable range of 'experience'" in order to transform it into a "more elaborate" device (Turing, c. 1951, p. 473). He called this simple machine a "child machine"—an "unorganised" machine that is to be organised in a manner analogous to the teaching of a human child (Turing, 1950, p. 460; 1948, p. 416).[12] Teaching a child, Turing said, "depends largely on a system of rewards and punishments" and accordingly his P-type child machine is organised by "pain" (or "punishment") and "pleasure" (or "reward") inputs (Turing, 1948, p. 425).[13] A pain signal cancels a tentative entry in the P-type's machine table and a pleasure signal confirms it. Organising the P-type requires additional inputs, since, Turing joked, if a child learns only by means of punishment and reward he or she "would probably feel very sore indeed" (1950, p. 461). Turing called these additional channels of communication "sense stimuli": "The sense stimuli are means by which the teacher communicates 'unemotionally' to the machine, i.e. otherwise than by pleasure and pain stimuli" (Turing, 1950, p. 461; 1948, p. 426). By implication, pain and pleasure signals are the mechanism of *emotional* communication.

Emotional communication is to facilitate machine learning. Turing said that the machine can be allowed to "wander at random through a sequence of situations …[with the experimenter] applying pain stimuli when the wrong choice is made, pleasure stimuli when the right one is made" (Turing, 1948, p. 428). "With appropriate stimuli on these lines … one may hope that … wrong behaviour will tend to become rare" (Turing, 1948, p. 425). The machine is to generalize from this "experience", as the human child does—Turing said, "At later stages in education the machine would recognize certain other conditions as desirable owing to their having been constantly associated in the past with pleasure, and likewise certain others as undesirable", with the result that the "schoolmaster" would not need to "apply the cane" any more (Turing, 1950, pp. 474-5).

Like Turing's other descriptions of his child machine, his use of the terms "pleasure" and "pain" is anthropomorphic; these expressions are used in the sense true of *human beings* (rather than to pick out some simple, proto-, or quasi-feeling hypothesized to be true of machines). But they are used only tongue-in-cheek: Turing no more suggested that his paper machine felt pain than that it was beaten with a schoolmaster's cane. In the human child pain is a feeling and Turing said that his definitions of punishment and reward signals do "not presuppose any feelings on the part of the machine" (Turing, 1950, p. 461). Instead, Turing used the words "pleasure" and "pain" as proxy for more mundane technical expressions. All talk of pain and pleasure can be cashed out without remainder—for example, we might refer to a pain signal merely as a *cancellation* instruction. "Pain" and "pleasure" are thus examples of the expressions that Drew McDermott called "wishful mnemonics" (see McDermott 1976; Proudfoot, 2011).

The current debate on the question whether machines can have emotions typically focuses on feelings or sentience and the consequent ethical issues for AI (on this debate see e.g. Arbib & Fellous, 2004; Adolphs, 2005; Parisi & Petrosino, 2010). Did Turing think that machines can have feelings? Certainly he criticized the "Argument from Consciousness"—the claim that, in Geoffrey Jefferson's words, "[n]o mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes [or] grief when its valves fuse" (Turing, 1950, p. 451). But Turing's criticism was only that this argument leads (he said) to solipsism; he did not claim that a machine *could* be conscious or sentient.[14] On the other hand, Turing also said, "It is customary … to offer a grain of comfort, in the form of a statement that some particularly human characteristic could never be imitated by a machine. … I cannot offer any such comfort, for I believe that no such bounds can be set" (Turing, 1951, p. 486). Although this is still short of the positive statement that machines *can* have feelings, it challenges the philosophical naysayer to present arguments against the idea of emotion in machines.

Turing's Lessons for Affective Computing Today

Turing's goal of building a child machine anticipated the modern aim to build a machine with the cognitive capacities of human infants, in particular infant-level social (or 'emotional') intelligence—a machine that learns in infant-like ways, autonomously, aided by human scaffolding and in a manner consistent with human biology and psychology. (For examples of this aim in developmental and social robotics, see e.g. Breazeal, 2009; Breazeal, Gray, & Berlin, 2009; Gold & Scassellati, 2007; Wu et al., 2009.) Turing's descriptions of his attempts to "educate" his P-type, A-type, and B-type unorganised machines contain fascinating insights for today's researchers (Turing, 1948, 1950).[15]

On the other hand, several of Turing's remarks on emotion and machines diverge from influential approaches in current AI. *First*, many researchers in affective computing embrace recent theories in cognitive neuroscience and evolutionary psychology that tie cognition tightly to social (or "emotional") intelligence. According to the canonical text, feelings "come first in development and retain a primacy that subtly pervades our mental life. … [F]eelings have a say on how the rest of the brain and cognition go about their business" (Damasio, 1994, pp. 159-60). According to a recent statement of the view, "emotional and rational processes are deeply intertwined, with each exerting major influences on the functioning of the other" (Levine & Perlovsky, 2010, p. 1). On this view, implementing emotions in a machine is essential, not only in practice for ease of human-machine interaction, but also in principle if we are to build an intelligent machine. Turing's remarks suggest a contrary view. In discussion with Turing, Max Newman spoke of "the reasoning side of thinking" (Turing, Braithwaite, Jefferson, & Newman, 1952, p. 498), implying that thinking is possible without affective states.[16] Turing's comments fit with Newman's. For example, Turing said that the "mysteries" of consciousness need not be solved in order to answer the question whether machines can think (Turing, 1950, pp. 452-3). He also spoke of "disabilities" that a machine may have that are "irrelevant" to the question whether it can think, and it seems that he regarded a machine's inability to "fall in love" or "enjoy strawberries and cream"—affective states—as irrelevant to intelligence (Turing, 1950, pp. 442, 453).

Second, several AI researchers aim to build hyper-realistic anthropomorphic robots (or software agents). Their motivation may be pragmatic, namely that humans can interact intuitively with "believable" machines or that human-inspired robotics helps us understand human minds, including consciousness. Or it may be philosophical, namely that human-like neurophysiology or physiognomy is essential to human-like cognition. In contrast, Turing hoped that "no great efforts will be put into making machines with the most distinctively human, but non-intellectual characteristics, such as the shape of the human body", since such machines "would have something like the unpleasant quality of artificial flowers" (Turing, 1951, p. 486).[17] His goal was a machine that had the "intellectual" rather than the "physical" capacities of a human being (Turing, 1950, p. 442). It seems that for Turing, even though "[o]ne way of setting about our task of building a 'thinking machine' would be to take a man as a whole and to try to replace all the parts of him by machinery" (Turing, 1948, p. 420), human-like neurophysiology and appearance are not necessary for human-like cognition.

Third, many researchers in emotional robotics claim that their machines (for example) *smile* or *frown* and have happy or sad expressions. Typically these researchers also qualify talk of a robot's "emotions", "goals", or "drives" by using square-quotes around these words,[18] to signal that they do not indicate *human* emotions; or they may say explicitly that their devices have merely simple or machine emotions. However, the claim that a robot or virtual character smiles or frowns, or has a happy or sad expression, implies that it has the relevant human emotion. Facial behaviour is a *smile* only if it has a certain meaning—the meaning that distinguishes a smile from a human grimace or facial tic, and from a chimpanzee's bared-teeth display. Likewise, a facial display is *happy* only if it signifies *happiness*. Ascribing emotional expressions or gestures to current social robots is unwitting anthropomorphism. In contrast, Turing's distinction between signal and affect—he said that his emotional signals to the P-type do "not presuppose any feelings on the part of the machine" (see above)—is one defence against unjustified anthropomorphism in AI.[19] In recent discussions, talk of "emotional" communication between human and machine tends to conflate the distinction between signal and affect. At best, modern talk of pleasure signals—or of the pleasure/displeasure dimension in machines—is ambiguous between mere reinforcement and actual feeling. At worst, the unreflective use of emotion-expression vocabulary settles prematurely the debate as to whether machines *can* have emotions.

Conclusion

Turing's remarks on emotion have been overlooked in the extensive literature on Turing and his test. Clarifying his uses of the term "emotional" is essential if we are to understand his answers to fundamental questions in the field of Artificial Intelligence.

Turing's notion of an *emotional concept* yields a new philosophical approach to the concept of intelligence and a distinctive argument against the computational theory of mind. Neglect of this notion has the result that Turing's philosophy of machine intelligence and test of intelligence in machines are misunderstood. According to the canonical behaviourist interpretation, Turing's test provides a criterion of thinking in terms of the capacity for or tendency to "thinking" or "intelligent" behaviour. (For influential examples see e.g. Block, 1981, 1995; French, 1996, 2000, 2007.) This reading is at odds with Turing's explicit claims about intelligence and his test's emphasis on the observer's reaction.[20] Turing's notion of *emotional communication* provides insights into building child machines and is also one barrier against unjustified anthropomorphism in AI.

Turing's remarks on emotion prompt many questions. Is intelligence a response-dependent property? Is cognition possible in the absence of affect? To build human-level social intelligence, should we first build human-like neurophysiology? These questions are crucial for affective computing today.

References

Adolphs, R. (2005). Could a Robot have Emotions? Theoretical Perspectives from Social Cognitive Neuroscience. In M. Arbib & J.-M. Fellous (eds), *Who Needs Emotions: The Brain meets the Robot*, pp. 9-28. Oxford University Press. http://www.cc.gatech.edu/~athomaz/papers/adolphs05.pdf.

Arbib, M.A. & Fellous, J-M. (2004). Emotions: from brain to robot. *TRENDS in Cognitive Sciences, 8(12)*, 554-61.

Block, N. (1981) Psychologism and Behaviorism. *Philosophical Review, 90(1),* 5-43.

Block, N. (1995) The Mind as the Software of the Brain. In E.E. Smith & D.N. Osherson (eds) *An Invitation to Cognitive Science, 2nd edition, Volume 3*, pp. 377-425. Cambridge, Mass: MIT Press.

Breazeal, C. (2009). Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of The Royal Society, Part B, 364,* 3527-38.

Breazeal, C., Gray, J., & Berlin, M. (2009). An Embodied Cognition Approach to Mindreading Skills for Socially Intelligent Robots. *The International Journal of Robotics Research, 28,* 656-80.

Brooks, R.A. Intelligence without Reason. (1995) In L. Steels & R.A. Brooks (eds), *The Artificial Life Route to Artificial Intelligence*, pp. 25-81. Hillsdale, NJ: Lawrence Erlbaum.

Caporael, L.R. (1986). Anthropomorphism and Mechanomorphism: Two Faces of the Human Machine. *Computers in Human Behavior, 2,* 215-234.

Churchland, P.M. (2008). On the Nature of Intelligence: Turing, Church, von Neumann, and the Brain. In Epstein, G. Roberts, & G. Beber, eds, *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Amsterdam: Springer.

Copeland. B.J. (2000). The Turing Test. *Minds and Machines, 10(4),* 519-39.

Copeland. B.J. (ed.) (2004). *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma*. Oxford University Press.

Copeland. B.J. & Proudfoot, D. (1996). On Alan Turing's Anticipation of Connectionism', *Synthese, 108,* 361-7. Reprinted in R. Chrisley (ed.) (2000). *Artificial Intelligence: Critical Concepts in Cognitive Science, Volume 2: Symbolic* AI. London: Routledge.

Copeland. B.J. & Proudfoot, D. (2008). Turing's Test: A Philosophical and Historical Guide. In R. Epstein, G. Roberts, & G. Beber, eds, *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Amsterdam: Springer.

Copeland. B.J. & Proudfoot, D. (2010). Deviant Encodings and Turing's analysis of computability. *Studies in History and Philosophy of Science Part A, 41(3*), 247-252.

Damasio, A. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: G.P. Putnam.

Duffy, B.R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems, 42,* 177-190.

Epstein, R., Roberts, G., & Beber, G., eds. (2008).*Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Amsterdam: Springer.

Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology, 59,* 255–78. doi: 10.1146/annurev.psych.59.103006.093629.

Evans, J. St. B. T. & Stanovich, K.E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science, 8,* 223-37.

Ford, K.M. & Hayes, P.J. (1998). On Computational Wings: Rethinking the Goals of Artificial Intelligence. *Scientific American Presents, 9(4)*.

French, R.M. (1996) Subcognition and the Limits of the Turing Test. In P. Millican & A. Clark (eds) *The Legacy of Alan Turing, Volume I: Machines and Thought*, pp. 11-26. Oxford University Press.

French, R.M. (2000) The Turing Test: the First 50 Years. *Trends in Cognitive Science, 4(3),* 115-22.

French, R. M. (2007). If it walks like a duck and quacks like a duck ... The Turing Test, Intelligence and Consciousness. In P. Wilken, T. Bayne, & A. Cleeremans (eds.). *Oxford Companion to Consciousness*, pp. 641-3. Oxford University Press.

Gold, K. & Scassellati, B. (2007). A Bayesian robot that distinguishes "Self" from "Other". *Proceedings of the 29th Annual Meeting of the Cognitive Science Society (CogSci2007)*. Nashville, Tennessee. http://cs-www.cs.yale.edu/homes/scaz/papers/Gold-CogSci-07.pdf.

Hanson, R. (1991). Has Penrose Disproved A.I.? *Foresight Update, 12,* 4-5. http://www.islandone.org/Foresight/Updates/Update12/Update12.2.html#anchor1339524. Accessed 12/03/2014.

Harnad, S. (2008). The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence. In R. Epstein, G. Roberts, & G. Beber, eds, *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Amsterdam: Springer. Latest version http://eprints.soton.ac.uk/262954.

Hayes, P.J. & Ford, K.M. (1995). Turing Test Considered Harmful. *IJCAI-95 Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, Quebec, August 20-25, vol. 1, Morgan Kaufman.

Kurzweil, R. (2002). Locked in his Chinese Room: Response to John Searle. In J. W. Richards (ed.), *Are We Spiritual Machines? Ray Kurzweil vs. the Critics of Strong A.I.* Discovery Institute, http://www.discovery.org.

Levine, D.S. & Perlovsky, L.I. (2010). Emotion in the Pursuit of Understanding. *International Journal of Synthetic Emotions, 1(2),* pp. 1-11.

Lucas, J.R. (1961). Minds, Machines, and Gödel. *Philosophy, 36,* 112-27.

Lucas, J.R. (1990). A paper read to the Turing Conference at Brighton on April 6th, 1990. http://users.ox.ac.uk/~jrlucas/Godel/brighton.html. Accessed 4/3/2014.

Lucas, J.R. (2008). Commentary on Turing's 'Computing Machinery and Intelligence'. In R. Epstein, G. Roberts, & G. Beber (eds), *Parsing the Turing Test: Philosophical and Methodological Issues*, pp. 67-70. Berlin: Springer.

Lucas, J.R. (2013). The face of freedom. Comment on "Consciousness in the universe. A review of the 'Orch OR' theory" by Stuart Hameroff and Roger Penrose. *Physics of Life Reviews, 431*. http://dx.doi.org/10.1016/j.plrev.2013.11.008.

McDermott, D. (1976). Artificial Intelligence Meets Natural Stupidity. SIGART Newsletter*, 57,* 4-9.

Moor, J.H. (2001) The Status and Future of the Turing Test. *Minds and Machines, 11,* 197-213.

Moor, J.H., ed. (2003). *The Turing Test: The Elusive Standard of Artificial Intelligence*. Dordrecht: Kluwer.

Parisi, D. & Petrosino, G. (2010). Robots that have emotions. *Adaptive Behavior, 18*, 453-69.

Proudfoot, D. (2005). A New Interpretation of the Turing Test, *The Rutherford Journal: The New Zealand Journal for the History and Philosophy of Science and Technology*, 1. rutherfordjournal.org.

Proudfoot, D. (2011). Anthropomorphism and AI: Turing's much misunderstood imitation game. *Artificial Intelligence, 175(5-6),* 950-7.

Proudfoot, D. (2013a). Rethinking Turing's Test. *Journal of Philosophy, 110(7),* 391-411.

Proudfoot, D. (2013b). Can a robot smile? Wittgenstein on facial expression. In T.P. Racine & K.L. Slaney (eds), *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology.* Basingstoke: Palgrave Macmillan.

Proudfoot, D. (forthcomingA). Turing's Child Machines. In Jonathan Bowen, Jack Copeland, Mark Sprevak, & Robin Wilson (eds), *The Turing Guide.* Oxford University Press.

Proudfoot, D. (forthcomingB). Turing's Concept of Intelligence. In Jonathan Bowen, Jack Copeland, Mark Sprevak, & Robin Wilson (eds), *The Turing Guide.* Oxford University Press.

Proudfoot, D. & Copeland, B.J. (2011). Artificial Intelligence. In E. Margolis, R. Samuels, & S.P. Stich, eds, *Oxford Handbook of Philosophy of Cognitive Science*, pp. 147-82. Oxford University Press.

Reddy, R. (2012a). Educating a Child Machine: The Unfinished Agenda of Human-Level AI. Talk at Rick Rashid's Festschrift, March 9, 2012. www.rr.cs.cmu.edu/EducatingCM.docx

Reddy, R. (2012b). Creating a Child Machine: Reflections on Turing's Proposal. ACM Turing Centenary Celebration Talk, June 15, 2012. www.rr.cs.cmu.edu.

Shah, H. & Warwick, K. (2010) Hidden Interlocutor Misidentification in Practical Turing Tests. *Minds and Machines, 20,* 441-54.

Turing, A.M. (1947) Lecture on the Automatic Computing Engine. Reproduced in Copeland (ed.) 2004; page references are to Copeland (ed.) 2004.

Turing, A.M. (1948) Intelligent Machinery. Reproduced in Copeland (ed.) 2004; page references are to Copeland (ed.) 2004.

Turing, A.M. (1950) Computing Machinery and Intelligence. Reproduced in Copeland (ed.) 2004; page references are to Copeland (ed.) 2004.

Turing, A.M. (c. 1951) Intelligent Machinery: a Heretical Theory. Reproduced in Copeland (ed.) 2004; page references are to Copeland (ed.) 2004.

Turing, A.M. (1951) Can Digital Computers Think? Reproduced in Copeland (ed.) 2004; page references are to Copeland (ed.) 2004.

Turing, A.M., Braithwaite, R., Jefferson, G., & Newman, M. (1952) Can Automatic Calculating Machines Be Said to Think? Reproduced in Copeland (ed.) 2004; page references are to Copeland (ed.) 2004.

Warwick, K. (2012). Not Another Look at the Turing Test! In M. Bieliková, G. Friedrich, G. Gottlob, S. Katzenbeisser, & G. Turán, eds, *BielSOFSEM 2012: Theory and Practice of Computer Science*. *Lecture Notes in Computer Science* 7147, pp. 130-40. Springer.

Warwick, K. & Shah, H. (2013) Good Machine Performance in Turing's Imitation Game. *IEEE Transactions on Computational Intelligence and AI in Games, 99*. http://ieeexplore.ieee.org.

Warwick, K., Shah, H., & Moor, J. (2013) Some Implications of a Sample of Practical Turing Tests. *Minds and Machines, 23,* 163-177.

Wu, Tingfan, Butko, Nicholas J., Ruvulo, Paul, Bartlett, Marian S., & Movellan, Javier R. (2009). Learning to Make Facial Expressions. *2009 IEEE 8th International Conference on Development and Learning*. http://mplab.ucsd.edu/wp-content/uploads/wu_icdl20091.pdf.

---

[1] This paper was written while I was a Gastprofessor at the Swiss Federal Institute of Technology (ETH), Zurich. I am most grateful to ETH, and to Professor Michael Hampe and Dr Giovanni Sommaruga, for this invitation. My thanks also to anonymous reviewers for the IJSE for their comments on an earlier draft of this paper.

[2] The argument in this section is set out in detail in Proudfoot, 2013a.

[3] Recent work on practical Turing tests reveals characteristic features of the machine and strategies of the interrogator that may increase the likelihood of interviewees' appearing intelligent in brief imitation games. See e.g. Shah & Warwick, 2010; Warwick & Shah, 2013; Warwick, Shah, & Moor, 2013.
[4] For a variety of objections to Turing's test, see e.g. Moor, ed., 2003; Epstein, Roberts, & Beber, eds, 2008. For assessment of influential objections to the test, see e.g. Moor, 2001; Harnad 2008; Proudfoot, 2011, 2013a; Copeland & Proudfoot, 2008, Proudfoot & Copeland, 2011.
[5] This question is distinct from Turing's *predictions* concerning how machines might fare in the imitation game. In his 1950 paper Turing said that "in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning" (Turing, 1950, p. 442). In his 1952 broadcast he made a stronger prediction, saying that it would be "at least 100 years" before a machine "stand[s] any chance [of passing the test] with no questions barred" (Turing, Braithwaite, Jefferson, & Newman, 1952, p. 495).
[6] Copeland points out this remark by Turing and discusses its significance (Copeland, 2004, p. 436). For a contrary view, which takes Turing's 1950 prediction to state the protocol for scoring the imitation game, see e.g. Warwick 2012.
[7] The Turing test has frequently been claimed to be subjective (e.g. Ford & Hayes, 1998, p. 79).
[8] At least arguably—there is significant debate on the objectivity of response-dependent concepts.
[9] Unless we embrace global response-dependence or take the view that encodings are observer-dependent (on the latter see Copeland & Proudfoot, 2010).
[10] According to Turing, whereas people who put forward objections such as the Argument from Consciousness would "mostly be willing to accept the imitation game as a basis for discussion", those who propose emotional objections "would probably would not be interested in any criteria" for a machine's having intelligence (Turing, 1950, p. 451).
[11] Lucas has argued this case in various papers. For Lucas's and his critics' core papers see the issue of *Etica & Politica* (Vol. 5, No. 1, 2003) on *John Lucas against Mechanism*; http://www2.units.it/etica/2003_1/index.html.
[12] On Turing's child machines and analogous recent work in AI, see Proudfoot 2013b, forthcomingA. For recent work in artificial intelligence that appeals explicitly to Turing's notion of a child machine, see e.g. Reddy, 2012a, 2012b.
[13] A P-type is 'an L.C.M. without a tape, and whose description is largely incomplete' (Turing 1948, p. 425).
[14] Likewise Turing's response to the "Argument from Various Disabilities" is that this objection is based only on extant machines (Turing, 1950, p. 453).
[15] A-type and B-type machines are networks of neuron-like elements; see Turing, 1948; Copeland & Proudfoot, 1996.
[16] Or that one "side" of thinking is possible without affect. This view implies a dual-process theory of thinking; for an overview see e.g. Evans 2008, Evans & Stanovich, 2013.
[17] Interestingly, this calls to mind critics of Turing's test who say that the test prioritizes contingent features of human intelligence over features of intelligence "shared by other species" (Hayes & Ford, 1995, p. 974).
[18] As did Turing when he first introduced the idea of using "pain" and "pleasure" stimuli to train machines (Turing, 1948, p. 425).

[19] Turing was well aware of (in his words) the "temptation to imagine intelligence" in machines and he said that interacting with even a paper machine gives "a definite feeling that one is pitting one's wits against something alive" (Turing, 1948, pp. 431, 412). Modern researchers, however, frequently succumb to anthropomorphism; on this see McDermott, 1976; Caporael, 1986; Duffy, 2003; Proudfoot, 2011.

[20] For criticism of behaviourist (and other) interpretations of Turing's test using Turing's notion of an emotional concept, see Proudfoot, 2013a; also Proudfoot 2005, forthcomingB.