

Disruptive Behaviors: AI, Robots and the Autism Spectrum Disorder

Gregory P. Garvey

Game Design & Development, Quinnipiac University
Connecticut, USA
greg.garvey@quinnipiac.edu

Abstract

Today we see the rise of new artificially intelligent entities. Some are embodied as robots and others as non-corporal AIs in devices, interfaces and games. Researchers in robotics and Artificial Intelligence and philosophers speculate that these entities will some day pass the Turing Test and exhibit Artificial Consciousness (AC) or Artificial General Intelligence (AGI), act as artificial moral agents (AMAs), be our lovers and even manifest the signs of experiencing pain and suffering. If such entities become our fellow workers, associates and companions shall these entities be extended the status of personhood with all the rights, privileges and protections under the law? How can we judge if they are truly conscious? Will the Turing Test be sufficient test to judge their fitness for citizenship? What if they exhibit behaviors that match the diagnostic criteria from the *American Psychiatric Association's* Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) for Autism Spectrum Disorder, Dissociative Identity Disorder or Antisocial Personality Disorder? This paper argues that a DSM "Turing Test" will be needed for what Marvin Minsky termed "self-improving" robots to determine if they are fit to work along side of and interact with human beings on a daily basis.

Keywords

Artificial Intelligence; Artificial Consciousness; Artificial General Intelligence; Artificial Moral Agents; Autism; Person; Rights; Dissociation; Psychopathology; DSM-V

Introduction

In 1979 Marvin Minsky cautioned that the first 'self-improving' robots and AIs may become psychotic and it will take "generations of theories and experiments to stabilize them." [1] Today we see the rise of new artificially intelligent entities. Some are embodied as robots and others as non-corporal AIs in devices, interfaces and games. Researchers in robotics and Artificial Intelligence and philosophers speculate that these entities will some day pass the Turing Test and exhibit Artificial Consciousness (AC) or Artificial General Intelligence (AGI), act as artificial moral agents (AMAs), be our lovers and even manifest the signs of experiencing pain and suffering. [2]

Robots Will Be People (too!)

Philosopher David Deutsch argues that Artificial General Intelligence (AGI) is not only possible but also such enti-

ties will indeed be self-aware and are most assuredly 'people'. [3] A report from the Future of Identity in the Information Society (FIDIS) concludes: "When it comes to attributing full legal personhood and 'posthuman' rights to new types of entities, the literature seems to agree that this only makes sense if these entities develop self-consciousness." [4]

If such entities become our fellow workers, associates and companions shall these entities be extended the status of personhood with all the rights, privileges and protections under the law? How can we judge if they are truly conscious? Will the Turing Test be sufficient test to judge their fitness for citizenship and/or Peter Singer's expanding circle for inclusion as a living being with rights? [5] What if Marvin Minsky is right and they do become psychotic? How do we diagnose their condition? What if they exhibit behaviors that match the diagnostic criteria for Autism Spectrum disorders?

A Freitas Jr. offers a relatively straightforward way to assess if AIs or robots are ready to be functioning members of society: "Certainly any self-aware robot that speaks English and is able to recognize moral alternatives, and thus make moral choices, should be considered a worthy "robot person" in our society. If that is so, shouldn't they also possess the rights and duties of all citizens?" [6].

David Deutsch suggests a test to verify if an entity qualifies as an AGI by determining whether or not "it lacked even a single cognitive ability that is characteristic of people."

In *Corpus Juris Roboticum*, Raymond August notes that Anglo-American Law has adopted the sanity test to judge competency. Under the Model Penal Code "insanity is the incapacity to either appreciate wrongfulness or conform to the requirements of the law." [7]

Presumably such a test could be given to a robot or AGI. However using rule based programming techniques, an AGI could easily follow the dictates of the law and "readily regurgitate statements of law on demand." August feels a sanity test alone is insufficient. Instead August proposes a more robust but simple test consisting of six questions:

1. Does it/he/she have a complex brain?
2. Is the brain capable of speculation, calculation and memory, in addition to the operation of subsystem or body parts?
3. Is the brain's capacity for speculation, calculation and memory comparable to that of a human?

4. Is the brain capable of learning, i.e., can it separate potentially useful information from useless information, and can it purge or discard the useless?
5. Is the brain's capacity to learn unlimited by subject matter, i.e., is it capable of invention?
6. Is the brain capable of using sensory devices to perceive its environment and to interface with humans even if those sensory devices are not connected?

If a robot passes the above test along with a sanity test then for August, “it seems logically, ethically and morally compelling not only to regard it as both human and sane, but also entitled to the rights of other “natural”, humans.

The Turing Triage Test

Another test that might be used to determine if we are willing to confer such rights to robots is proposed by Robert Sparrow. Alan Turing first proposed the Turing Test in 1950 and referred to it as the “imitation game.” [8] Sparrow extends Turing’s imitation game to what he calls the Turing Triage Test. [9] Like the trolley problem from Game Theory, the Turing Triage Test posits that two lives are at stake and only one can be saved. He continues: “We will know that machines have achieved moral standing comparable to a human when the replacement of one of the patients with an artificial intelligence leaves the character of the dilemma intact. That is, when we might sometimes judge that it is reasonable to preserve the continuing existence of the machine over the life of the human being. This is the ‘Turing Triage Test’.”

Cyborg Citizen Turing Test

Chris Hables Gray also references the Turing Test (i.e. the imitation game) and proposes a ‘double-blind’ Cyborg Citizen Turing test “to see which entities can actually operate in our discourse community, and which cannot.” [10] Gray sees the ability to fully participate in the discourse of citizenship as judged by jury of peers (other citizens) as the measure of inclusion under the protections of a Bill of Rights for robots, AIs and their fellow travelers.

The Replicant Test

The Replicant Test was featured in the sci-fi classic Blade Runner based on Philip K. Dick’s novel “Do Androids Dream of Electric Sheep?” [11] In the film, Harrison Ford plays Richard Deckard, a “Blade Runner” whose job is to hunt down artificial humanoids known as replicants. Once he finds them he terminates them. To determine if a subject is human or replicant the Blade Runner Deckard uses a Voight-Kampff machine. [12] This is a kind of lie detector machine that measures “respiration, blush response, heart rate, and eye movement in response to emotionally provocative questions” [13] while the Blade Runner asks the subject 20 to 30 cross-referenced questions.

An updated version of the replicant test can be found on the OK Cupid dating site to determine if the online object

of your affection is a replicant. [14] A sample question seek to probe the replicant’s life history or lack of one:

From the choices below, Choose only the good things that come into your mind. About your mother...

- Kind, loving, a great cook
- Strict, intimidating, kind of loud
- sweet, smart, a nice person
- I never knew my mother

These questions point to an actual lived life. “Robots” or replicants have fabricated life histories. Careful questioning would begin to reveal inconsistencies in this “imitation game.”

Eliza

This example from this questionnaire recalls Joseph Weizenbaum’s Eliza Program. [15] This software demonstration used a text based interface to simulate Rogerian psychotherapy by asking leading questions such as “Tell me more about your mother....”. That program famously had no understanding of context or meaning and yet Weizenbaum writes that people took it quite seriously and responded to the questions. This points to our human gullibility and desire to project agency where there is none.

Autism Spectrum Disorders

However it is quite likely that humanoid robots and other species of AIs will tumble into the uncanny valley [16]. As humanoid robots appear and act more life-like we will notice the differences and will recoil or feel disgust. While there has been a rapid advances in the AI labs around the world in a race to develop social robots, techniques of “affective” computing that can read human emotions and display appropriate emotions is still plagued by obvious shortcomings.

Humanoid robots will betray serious limitations especially in social situations in very obvious and disruptive ways. These shortcomings will likely manifest as inappropriate verbal responses, abnormal eye contact and repetitive actions. Since robots will have simulated emotional states and will depend on robust facial recognition software “empathy errors” will be prevalent. Humans can easily detect inconsistencies. Robots will likely have fictionalized life histories, and will lack true subjective experiences (qualia). Yet they will likely be programmed to report that they do see the color “red” or feel love or have memories of a childhood or even their parents. However it may be beside the point if a robot “experiences” internal states.

Cynthia Breazeal’s doll like Kismet [17] to David Hansen’s hyper-real bots [18] robots display communication deficits that induce the uncanny valley and are uncannily similar to a range of autistic behaviors.

Social Communication Disorder

The *American Psychiatric Association’s* Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders

identifies the following diagnostic criteria under Social (Pragmatic) Communication Disorder (code as item 315.39 in the DSM-V): [19]

So the question becomes will humanoid robots have “persistent difficulties in the social use of verbal and non-verbal communication as manifested by all of the following:”

1. Deficits in using communication for social purposes, such as greeting and sharing information, in a manner that is appropriate for the social context.
2. Impairment of the ability to change communication to match context or the needs of the listener, such as speaking differently in a classroom than on the playground, talking differently to a child than to an adult, and avoiding use of overly formal language.
3. Difficulties following rules for conversation and storytelling, such as taking turns in conversation, rephrasing when misunderstood, and knowing how to use verbal and nonverbal signals to regulate interaction.
4. Difficulties understanding what is not explicitly stated (e.g., making inferences) and nonliteral or ambiguous meanings of language (e.g., idioms, humor, metaphors, multiple meanings that depend on the context for interpretation).

Will these deficits of humanoid robots “result in functional limitations in effective communication, social participation, social relationships, ...or occupational performance, individually or in combination”?

Aspergers

Aspergers was eliminated as a category in the DSM-V and is simplified as the Autism Spectrum Disorder (coded as item 299.00 in the DSM-V) which has the following diagnostic criteria:

Will humanoid robots have “persistent deficits in social communication and social interaction across multiple contexts, as manifested by the following:”

1. Deficits in social-emotional reciprocity, ranging, for example, from abnormal social approach and failure of normal back-and-forth conversation; to reduced sharing of interests, emotions, or affect; to failure to initiate or respond to social interactions.
2. Deficits in nonverbal communicative behaviors used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expressions and nonverbal communication.
3. Deficits in developing, maintaining, and understanding relationships, ranging, for example, from difficulties adjusting behavior to suit various social contexts; to difficulties in sharing imaginative

play or in making friends; to absence of interest in peers.

4. Difficulties understanding what is not explicitly stated (e.g., making inferences) and nonliteral or ambiguous meanings of language (e.g., idioms, humor, metaphors, multiple meanings that depend on the context for interpretation).

While space does not permit a more detailed discussion of each of the above diagnostic criteria, this author submits that by there are no robots in existence today that would escape the diagnostic “net” of the above criteria. However, this author does also make the claim that we will need to systematize the above criteria into a kind of new Turing Test for Autism. We will want to know how well our synthetic hardware or wetware sisters and brothers will do in social situations. Will they embarrass us? Will they offer us comfort and be supportive? And what of more dystopian futures that are a replicant meme of Hollywood movies i.e. the malevolent robot or evil AI bent on world domination.

Anti-social Personality Disorder

Will “self-improving” robots develop on their own new algorithmic solutions to classic examples of game theory such as the “Prisoner’s Dilemma”? [20] Will the robot reach the inference that in every social situation it is a zero sum game and it must select the win state at all costs and ignore Isaac Asimov’s Three Laws for Robots? [21]

We can again look to the DSM-V for the diagnostic criteria to determine if our robots, our sexbots, robot slave workers and their fellow travelers are psychotic or worse psychopaths! The DSM-V now calls psychopathology “anti-social personality disorder.” The primary diagnostic criteria are as follows:

1. Impairments in self-functioning (a or b):
 - a. Identity: Ego-centrism; self-esteem derived from personal gain, power, or pleasure.
 - b. Self-direction: Goal-setting based on personal gratification; absence of prosocial internal standards associated with failure to conform to lawful or culturally normative ethical behavior.

AND

2. Impairments in interpersonal functioning (a or b):
 - a. Empathy: Lack of concern for feelings, needs, or suffering of others; lack of remorse after hurting or mistreating another.
 - b. Intimacy; Incapacity for mutually intimate relationships, as exploitation is a primary means of relating to others, including by deceit and coercion; use of dominance or intimidation to control others.

Will robots present pathological personality traits in the following domains:

1. Antagonism, characterized by:
 - a. Manipulativeness: Frequent use of subterfuge to influence or control others; use of seduction, charm, glibness, or ingratiation to achieve ones ends.
 - b. Deceitfulness: Dishonesty and fraudulence;

misrepresentation of self; embellishment or fabrication when relating events.

c. Callousness: Lack of concern for feelings or problems of others; lack of guilt or remorse about the negative or harmful effects of ones actions on others; aggression; sadism.

2. Disinhibition, characterized by:

a. Irresponsibility: Disregard for – and failure to honor – financial and other obligations or commitments; lack of respect for – and lack of follow through on – agreements and promises.

b. Impulsivity: Acting on the spur of the moment in response to immediate stimuli; acting on a momentary basis without a plan or consideration of outcomes; difficulty establishing and following plans.

c. Risk taking: Engagement in dangerous, risky, and potentially self-damaging activities, unnecessarily and without regard for consequences; boredom proneness and thoughtless initiation of activities to counter boredom; lack of concern for ones limitations and denial of the reality of personal danger.

Although many of these traits appear to be uniquely human (e.g. boredom) such behavior by a robot would be interpreted by humans as fitting the DSM descriptions.

Discussion and Conclusions

In recent years AI has made impressive strides performing tasks considered difficult for humans. The developers of Google's Deep Mind project reported in Nature [22] that the "deep Q-network agent" using reinforcement learning, received "only the pixels and the game score as inputs, was able to surpass the performance of all previous algorithms and achieve a level comparable to that of a professional human games tester across a set of 49 games..." Ginni Rometty, the Chairman and CEO of IBM claims that as Watson, the AI software that beat human Jeopardy winners, "gets smarter, his ability to reason is going to exponentially increase," and "In the future, every decision that mankind makes is going to be informed by a cognitive system like Watson," she said, "and our lives will be better for it." [23]

References

1. Marvin Minsky, "Computer Science and the Representation of Knowledge," in *The Computer Age: A Twenty-Year View*, ed. by Michael Dertouzos and Joel Moses (Cambridge and London: MIT Press 1979), 392–421.
2. Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (Oxford and New York: Oxford University Press 2009), 4.

Google's Deep Mind and IBM's Watson which both employ using powerful learning algorithms could arguably be described as "self-improving". Researchers do not consider these "AIs" as conscious and in fact, find the question of consciousness irrelevant to their work. However, as robots are introduced into our lives as fellow workers, associates and companions they will possess similar computation power and capacity to learn and "self-improve."

Researcher Julie Carpenter interviewed military Explosive Ordnance Disposal personnel who use robots for bomb disposal. Her research that "patterns in their responses indicated they sometimes interacted with the robots in ways similar to a human or pet," [24] If soldiers in the field form emotional attachments to robots what will happen when we work along side of humanoid robots and even have them as surrogate companions or lovers?

Even with behavioral deficits perhaps our future companion robots will be "just good enough" to pass the Turing Test of the imitation game. However, such robots with their access to vast data coupled with learning algorithms could become unpredictable. What if the robot lover learns or develops an algorithm that trumps conventional notions of moral behavior and begins to show anti-social tendencies or a "lack of concern for the concerns or feelings of others"? What if the robot is infected by malware or is hacked, lapses into silence, displays repetitive activities akin to autism or acts as if it has a new identity? What if there are accumulated errors and our fellow worker robots reports seeing or hearing things that are not there?

Will the Consumer Protection Agency of the future require robot workers and robots companions be continually certified to possess "normal" psychological profiles in order to protect end users? Will competition in the marketplace drive manufacturers and retailers seek to advertise that their robots pass the DSM Turing Test?

Acknowledgements

I am grateful for the inspiration provided by the speaker series sponsored by the Yale Interdisciplinary Center for Bioethics Technology and Ethics Study Group.

3. David Deutsch, "Philosophy will be the key that unlocks artificial intelligence," *The Guardian*, October 3, 2012, accessed December 30, 2014, <http://www.theguardian.com/science/2012/oct/03/philosophy-artificial-intelligence>
4. WP17, "D17.3: Bridging the accountability gap: rights for new entities in the information society?" ed. Bert-Jaap Koops, Mireille Hildebrandt and David-Oliver Jaquet-Chiffelle, FIDIS Future of Identity in the Information Society, accessed December 30, 2014, http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp17-del17.3-rights_for_new_entities_def.pdf

5. Peter Singer, *Animal Liberation: A New Ethics for our Treatment of Animals* (New York: Random House, 1975).
6. Robert A. Freitas Jr., "The Legal Rights of Robots," *Student Lawyer* 13, (January 1985): 54-56, accessed December 30, 2014, <http://www.rfreitas.com/Astro/LegalRightsOfRobots.htm>
7. Raymond August, "Corpus Juris Roboticum," *Computer Law Journal* 8 (1988): 375.
8. Alan Turing, "Computing Machinery and Intelligence," in *The New Media Reader*, ed. Noah Wardip-Fruin and Nick Montfort (Cambridge and London: MIT Press, 2003), 50.
9. Robert Sparrow, "The Turing Triage Test," *Ethics and Information Technology* Vol. 6 No. 4 (2004): 203-213.
10. Chris Hables Gray, "The Ethics and Politics of Cyborg Embodiment: Citizenship as a Hypervalue," *Cultural Values* Vol. 1 No. 2 (1997): 252-258.
11. Philip K. Dick, *Do Androids Dream of Electric Sheep?* (New York: Doubleday, 1968).
12. Wikipedia, s.v. "Blade Runner", accessed January 12, 2015, http://en.wikipedia.org/wiki/Blade_Runner
13. Paul M. Sammon, *Future Noir: the Making of Blade Runner*, (London: Orion Media, 1996), 106-107.
14. "The BladeRunner Replicant Test", accessed January 12, 2015, <http://www.okcupid.com/tests/the-bladerunner-replicant-test>
15. Joseph Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, Vol. 9 No. 1 (1966), 36-45.
16. Wikipedia, s.v. "The Uncanny Valley", accessed January 12, 2015 http://en.wikipedia.org/wiki/Uncanny_valley
17. "Kismet (Robot)", accessed January 12, 2015 http://en.wikipedia.org/wiki/Kismet_%28robot%29
18. Hanson Robotics, accessed January 12, 2015 <https://hansonrobotics.wordpress.com/>
19. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-V*, (Washington, DC: American Psychiatric Association 2013), 299.
20. Wikipedia, s.v. "The Prisoner's Dilemma", accessed May 30, 2015, http://en.wikipedia.org/wiki/Prisoner's_dilemma
21. Wikipedia, s.v. "The Three Laws of Robotics", accessed May 30, 2015, http://en.wikipedia.org/wiki/Three_Laws_of_Robotics
22. Volodymyr Mnih, Koray Kavukcuoglu and David Silver, "Human-level control through deep reinforcement learning," *Nature* 518, (2015): 529-533.
23. Lauren F. Friedman, "The CEO of IBM just made a jaw-dropping prediction about the future of artificial intelligence," *Business Insider Australia*, May 15, 2015, accessed May 30, 2015, <http://www.businessinsider.com.au/ginni-rometty-on-ibm-watson-and-ai-2015-5>
24. Doree Armstrong, "Emotional attachment to robots could affect outcome on battlefield," University of Washington website, accessed May 30, 2015,

<http://www.washington.edu/news/2013/09/17/emotional-attachment-to-robots-could-affect-outcome-on-battlefield/>

Bibliography

- Deutsch, David , "Philosophy will be the key that unlocks artificial Intelligence," *The Guardian*, October 3, 2012, <http://www.theguardian.com/science/2012/oct/03/philosophy-artificial-intelligence>.
- Dick, Philip K., "Do Androids Dream of Electric Sheep?" (New York: Doubleday, 1968).
- Freitas Jr., 5. Robert A., "The Legal Rights of Robots," *Student Lawyer* 13, (January 1985): 54-56, accessed December 30, 2014, <http://www.rfreitas.com/Astro/LegalRightsOfRobots.htm>
- Gray, Chris Hables, "The Ethics and Politics of Cyborg Embodiment: Citizenship as a Hypervalue," *Cultural Values* 1 no. 2 (1997): 252-258.
- Koops, Bert-Jaap. ed., Mireille Hildebrandt and David-Oliver Jaquet-Chiffelle, WP17, "D17.3: Bridging the accountability gap: rights for new entities in the information society?"
- Minsky, Marvin, "Computer Science and the Representation of Knowledge," in *The Computer Age: A Twenty-Year View*, ed by Michael Dertouzos and Joel Moses, (Cambridge and London: MIT Press 1979), 392-4214.
- Sammon, Paul M., *Future Noir: the Making of Blade Runner*, (London: Orion Media, 1996), 106-107.
- Peter Singer, *Animal Liberation: A New Ethics for our Treatment of Animals* (New York: Random House, 1975).
- Sparrow, Robert, "The Turing Triage Test," *Ethics and Information Technology* 6 no. 4 (2004): 203-213.
- Alan Turing, "Computing Machinery and Intelligence," in *The New Media Reader*, ed. Noah Wardip-Fruin and Nick Montfort (Cambridge and London: MIT Press, 2003), 50.
- Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong*(Oxford and New York: Oxford University Press 2009), 4.
- Weizenbaum, Joseph, "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, 9 no. 1 (1966), 36-45.

Author Biography

Gregory P. Garvey, MFA, MSVS, BS is the Chair of the Department of Visual and Performing Arts and also serves as Director of the Game Design and Development Program. Previously at Quinnipiac University he was the Visiting Fellow in the Arts and also was an Associate Artist of the Digital Media Center for the Arts at Yale University. Prior to joining Quinnipiac University he was Chair of the Department of Design Art at Concordia University in Montreal and was a member of the Board of Directors of the Montreal Design Institute. From 1983-85 he was a Fellow at the Center for Advanced Visual Studies at MIT.