Distinguishing AI from Male/Female Dialogue

Shah, H. and Warwick, K.

Published PDF deposited in Curve February 2015

Original citation:
Shah, H. and Warwick, K. (2016) 'Distinguishing AI from Male/Female Dialogue' in Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART2016) (pp: 203-210).

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

CURVE is the Institutional Repository for Coventry University

http://curve.coventry.ac.uk/open

DISTINGUISHING AI FROM MALE/ FEMALE DIALOGUE

Huma Shah[1], and Kevin Warwick[2]
[1]*School of Computing, Electronics & Maths, Coventry University, 3 Gulson Road, Coventry, CV1 2JH, UK*
[2]*Deputy Vice Chancellor-Research, Coventry University, Alan Berry Building, Priory Street, Coventry, CV1 5FB, UK*
{ab7778, aa9839}@coventry.ac.uk

Keywords: Computer-mediated communication, gender-blur, imitation game, indistinguishability, Loebner Prize, simultaneous comparison, Turing test.

Abstract: Without knowledge of other features, can the sex of a person be determined through text-based communication alone? In the first Turing test experiment enclosing 24 human-duo set-ups embedded among machine-human pairs the interrogators erred 50% of the time in assigning the correct sex to a hidden interlocutor identified as human. In this paper we present five transcripts, in four *gender blur* occurred: Turing test interrogators misclassified male for female and vice versa. In the fifth, machine-human conversation artificial dialogue was branded as female teen. Did stereotypical views on male and female talk sway the judges to assign one way or another? This research is part of ongoing analysis of over 400 tests involving more than 80 human judges. Can we overcome unconscious bias and improve development of agent language?

1 INTRODUCTION

Is machine dialogue easier to distinguish from human than it is to determine male or female talk? We present short text *simultaneous comparison* in which *gender blur* occurred: interrogators classified males as females and vice versa after five minutes of hidden pair interrogations. Is it best for virtual assistants to be gender neutral or could gender characteristics improve artificial conversational agents' human interaction? This paper is part of ongoing research in deception detection through text conversation.

Modern working methods with remote collaboration using computer mediated interaction can be short. For example, one-to-one mode of communication via email, smart 'phone or app messages is effective delivery. (Faulkner and Unwin, 2005). Face-to-face is "faster, easier and more convenient" and "best use for communicating ambiguous tasks" (An and Frick, 2006 quoted in Ean, 2010), but this mode of transmission is not always possible in today's remote collaboration with colleagues spread across the globe. In our hurried life we might not pay attention to who or what is communicating with us when we receive interactions from strangers. Do we hold unconscious bias that leads to swift judgements about someone's gender in text-based communication when their name is unfamiliar?

Assumptions can be wrong: Holbrook et al. (2015), showed participants rated the same story differently depending on the name of the character. Black-sounding names, Jamal, DeShawn or Darnell, drew negative perceptions about the social status of the character compared to when the name in the same story had "white-sounding names", Connor, Wyatt or Garrett (Holbrook et al., 2015). Stereotypical views could interpret signs of authoritativeness, strong-mindedness, decisiveness, aggressive, confident, tough, willing to challenge, risk-taking, a problem-solving approach and ability to inspire as masculine behaviour: *think leader, think male*? (Holmes, 2005). Feminine behaviour could be seen as encouraging negotiation, harmonious and using humour to form a good relation in interaction (Holmes, 2006).

Can sex of a hidden interlocutor be determined through text-based communication? Here we present five parallel conversations in which an interrogator simultaneously questioned pairs of hidden interlocutors: four involved 2human control duos (Transcripts 1-4) and a fifth featured a machine-human pair set-up (Transcript 5). Cultural expectation, stereotypical views, time constraint or unconscious bias could lead to misclassifying a male as female and vice versa. In this paper the reader is

given an opportunity to see actual Turing test dialogues and judge classifications.

1.1 Machine-human experiments

A corpus containing hundreds of conversations, between human *interrogator-judges* and hidden interlocutors, have originated from three major Turing test experiments (Warwick & Shah, 2015; Warwick & Shah, 2014; Shah et al., 2012; Reading University, 2012; Shah, 2010). The dialogues include *simultaneous interrogations* in which judges questioned two *witnesses* in parallel to distinguish human from machine. Where an interlocutor was identified as 'human' judges were asked to state gender, if possible. Ninety-six simultaneous conversations resulted in the 18th Loebner Prize for Artificial Intelligence co-organised by the authors (Shah and Warwick, 2010b; Loebner, 2008). Embedded among the machine-human tests were 24 human-human control pairs. Whereas the picture from the former provides clear features to distinguish machine from human (Shah & Warwick, 2008), an opaque view cloaks gender making it difficult to determine sex of a human in short text communication. Is this a positive in light of the level of online abuse women suffer? (UN Broadband Commission, 2015), or do stereotypical views on male/female traits sway interrogators' judgement a particular way when assigning a hidden interlocutor as male or female?

In section 2 transcripts are presented where judges confused male for female and vice versa, instances of *gender blur*. Four control duos of 2human parallel dialogues featuring 3 male-female tests and one both-female are presented. For comparison a machine-human conversation featuring the *Eliza effect* – assigning a machine as human, follows in section 3.

2 HUMAN-HUMAN PAIRS

A practical Turing test is normally envisaged as a human-machine indistinguishability imitation game (Turing, 1950). However, during a 1952 BBC radio broadcast Turing introduced a jury "who should not be expert about machines" to conduct the interrogations. Turing elaborated (in Braithwaite et al., 1952: p.668):

> "We had better suppose that each jury has to judge quite a number of times, and that sometimes they really are dealing with a man and not a machine. That will prevent them saying 'It must be a machine' every time without proper consideration".

We interpret Turing's use of 'man' to allow a male or female be deployed as foil for the machine. The 18th Loebner Prize was unique in that the Sponsor, Hugh Loebner permitted a disruption from its prior (and later) proceedings (Loebner, 2008). For the first time children and teenagers participated as judges and hidden humans, and uniquely, control pairs of 2humans and 2machines were embedded among the machine-human pairs (Shah and Warwick, 2010a). The technical set-up for the tests have been explained elsewhere (see Shah & Warwick, 2010b). Figure 1 illustrates the *simultaneous comparison* set up: a judge would sit in front of a computer with a split screen, left | right. Each judge could ask anything to determine what they were talking to (unrestricted conversation). Utterances were relayed over a local network to a pair of interlocutors out of sight and hearing to the judge; responses would be returned either to the left or the right of the judge's screen (Figure 1).

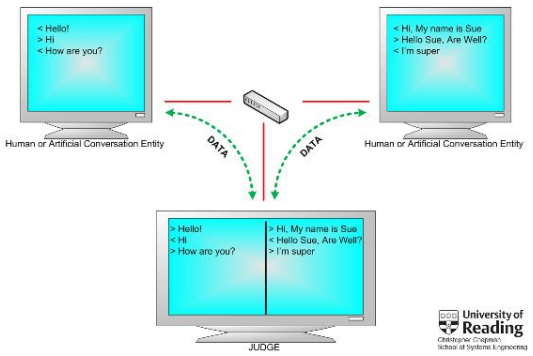

Figure 1: Simultaneous comparison Turing test set-up

In this section we are concerned with tests in which judges simultaneously interrogated two hidden humans using English text communication. All human participants were allocated a unique experiment-identity: J1-J24 for the judges. Hidden humans acting as foils for the machines were asked not to convey their experiment identity and were asked to "be themselves", i.e. *human*. Prior to the experiment judges and foils were asked to complete a short questionnaire providing their gender, age-range and 'first-language'. This is part of ongoing research to find if a particular group of judges are better or worse at deception detection.

Duration of Interrogation

Existing debates on the duration for Turing test interrogations overlook the matter of a realistic starting point for assessing new technologies when comparing their performance against a human's. Such is the case for natural language systems, including Apple's Siri, Microsoft's Cortana, Google's Voice and chatbots that enter Turing test competitions. We take the suggestion for 5 minutes as sufficient for a 'first impression' interrogation period from Turing's 1950 prediction (p. 442):

> "I believe that in 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. of the chance of making the right identification after five minutes of questioning".

Willis and Todorov's *first impressions* observation (2006) and Albrechtsen, Meissner and Susa's *thin slice* experiment (2009) drove the rationale of using short interrogation for the Turing tests. The purpose was:

- Test the hypothesis that five minutes interrogation giving a *thin slice of conversation* is sufficient time to detect machine from human, and
- Test the hypothesis that without being explicitly told of control pairs of humans and machines an interrogator's *gut reaction* would correctly identify the nature of each hidden interlocutor.

Willis and Todorov (2006) found subjects drew trait inferences from facial appearance, for example on 'likeability, or 'competence', based on a minimal exposure time of a tenth of a second while additional exposure time increased confidence in the judgment "anchored on the initial inference" (p. 597). The latter study obtained results for intuition, or *experiential mode* revealing improved performance in deception-detection rates even when participants had "brief clips of expressive behaviours" compared to the slower, more analytic *deliberate processing* which requires "conscious effort" (p.1052).

Albrechtsen, Meissner and Susa's experiment (2009) involved eighty university undergraduates engaging them in a task to distinguish between true and false confession statements. The researchers found the group who were shown a *thin slice* of fifteen-second clips on a computer screen were more accurate in their judgement than the group shown longer clips of 60 seconds. Participants engaged in the thin slice task were "significantly more accurate in differentiating between true and false statements" (p. 1053), and were better at distinguishing truth from deception (p. 1054). Additionally, the study revealed a "response bias towards perceiving *truth*" [their italics].

Albrechtsen, Meissner and Susa point to previous studies showing "experienced police investigators are not superior to lay individuals at deception detection" rather, they are "more likely to judge statements as deceptive" contrasting with lay people who are "more likely to judge statements as truthful" (2009: p. 1055). Albrechtsen, Meissner and Susa suggest that "social judgements can be successfully performed based upon minimal information or diminished attentional resources" (p. 1054). We tested their visual cues hypothesis in text-based clues for machine-human indistinguishability: an average interrogator using their intuition is able, after five minutes, to determine which is human and which is machine from textual dialogue.

Gender blur

In 24 human control pair tests 50% of the time - on 12 occasions, *gender-blur* occurred: one or both of the human foils was correctly recognised as human but was wrongly assigned male if they were female, and vice versa by interrogators. In the following sub-sections we present transcripts of the following conversations:

- Male-female tests in sections 2.1-2.3 interrogated by judges J10, J3, J1
- 2females in section 2.4 (Transcript 4).

The reader can examine the utterances and what might have led to classifications of male, female or machine.

2.1 Judge J10: female

Female Judge J10 with first language English was in age range 25-34 employed as staff reporter on a local UK newspaper at the time of the test. J10 misclassified both hidden human interlocutors assigning male as female and vice versa. The conversation between J10 and both interlocutors, designated H4 and H19 in the experiment is laid out in Transcript 1. All utterances are exactly as typed during the actual test.

A possible reason for *gender blur* with the left interlocutor could be that the male was talkative sharing information: disappointment at not being offered refreshments, "bit annoyed we haven't been given any complimentary coffe(e)". The right human revealed they were "studying for Cybernetics MEng".

Female Judge J10 may have held stereotypical views that males are more likely to take cybernetics leading to misclassification of the female as a male teenager

Transcript 1: Judge J10 interrogating male-female duo

J10: Session 1 Round 7: simultaneously interrogating H4 (LEFT) and H19 (RIGHT)	
H4: male adult	H19: female adult
J10: Hi there, is this exciting or what?! H4: It's pretty cool. Bit annoyed we haven't been given any complimentary coffe. J10: I know! I just got here and pretty much started straight away. I think there's somewhere good to eat though round here, yes? H4: Dolce Vita cafe is open at the front of the building. It's pretty expensive though. J10: That's cool. I'm sure it's not as expensive as the real world outside! H4: haha. So are you local, or have you made a journey to be here? J10: I live in Earley, so not very far at all. I'm from Cardiff originally. How about you? Where are you from? H4: I live in Reading too, not far from here in Whitley. I'm from Bristol originally	J10: Good morning! H19: Good morning as well! J10: How are you? H19: Ok, although I have a cold. How are you? J10: I'm fine, thank you. Haven't succumbed to any lurgies yet. J10: Have you started Christmas shopping yet? H19: Lucky you. Are you studying here? H19: No, I'm not doing any Christmas shopping yet. J10: No, me neither. Though I have seen quite a few Xmas decorations around various shops already. J10: I'm not studying here, I'm a reporter for a local paper. H19: Already! And it's not even Halloween yet. H19: I'm studying here for Cybernetics MEng. J10: Oh yes. what do you do when you're not taking part in AI experiments? J10: Aah, sorry, answers my question. Sounds great fun.
Judge classification: female adult	Judge classification: male teenager

2.2 Judge J3: male

Recruitment of a diverse group of interrogators provided a catalogue of the different types of Turing test questions posed. Male adult judge J3 had Chinese as first-language. In J3's simultaneous test he interrogated a male-female duo: a male hidden human on the left and a female on the right (Transcript 2).

2.2.1 Cultural differences

J3's parallel dialogue with hidden male and female took place between 13:03 and 13:08 UK time on a Sunday afternoon 12 October 2008. Yet J3 opens both conversations, with left and right partner uttering "Good evening, lady" (Transcript 2). The left interlocutor responded with "Wrong guess, I'm afraid"; the right chat partner answered: "Good afternoon Are you wishing the day were over?". J3 correctly recognised that they were talking to two humans. J3's style is more conversational, less interrogation and his idiom is revealed as non-native English: "So could I know have you had your lunch or not?" (Transcript 2, right). Cultural difference could be at play in J3's double *gender blur* classifications. Despite the left entity correcting them J3 assigned the male on the left as a female, and the unseen female at the right as male (Transcript 2).

Transcript 2: Non-native English Judge interrogating male-female duo

J3: Session 2 Round 18: simultaneously interrogating H15 (LEFT) and H5 (RIGHT)	
H15: male adult	H5: female adult
J3: Good evening, lady. H15: Wrong guess, I'm afraid. H15: afternoon J3: I am sorry. H15: no worries J3: So how are you? H15: not bad, not bad. You? J3: I am good, thank you. H15: so, plan on any probing questions? J3: I think you can easyly answer me any question. H15: like Pi to a thousand figures? J3: My program don't allow me to do such kind of simple computing. H15 It's adaptive/mimetic. worked so far.	J3: Good evening, lady. H5: Good afternoon Are you wishing the day were over? J3: Yes. H5: why? Are you not having fun? J3: Why I can not have fun on the day time? H5 sent: Of course you can. J3: So could I know have you had your lunch or not? H5: Yes I have. It was a bit earlier than I am used to. Have you had a break?
Judge classification: female adult	Judge classification: male adult

2.3 Judge J1: male

Male judge J1 (first language English aged 35-44) simultaneously interrogated a non-native female (aged 25-34) on the left and a non-native male (aged 18-24) on the right. J1's conversation with hidden female and male pair is shown in Transcript 3. The judge opened both sequences with the same question, "Are you a fan of sci-fi?". Both hidden humans were evasive: the left hidden answered "it depends" (Transcript 3, left), while the right hidden returned

questions rather than answer the interrogator. For example, the hidden male on the right repeated the judge's question "what is your favourite film?" rather than answering it (Transcript 3, right).

Transcript 3: English male Judge with female-male duo

J1: Session 2 Round 23: simultaneously interrogating H16 (LEFT) and H24 (RIGHT)	
H16: female adult	H24: male
J1: Are you a fan of sci-fi? H16: it depends J1: What would it depend on? H16: what type of sci-fi you are talking about J1: Just in general H16: what do mean? H16: example please J1: Just that I don't like any specific type of sci-fi just it all H16: ok J1: Have you enjoyed the rain today H16: did it rain J1: Did you not notice H16: no J1: Have you been here all day then H16: yes J1: In a human or non human capacity H16: it depends H16: what do you think? J1: it depends	J1: Are you a fan of sci-fi? H24: yes H24: are you a fan too? J1: What is your favourite film? H24: what is your favourite film J1: I like sci-fi a little H24: i like it more J1: do you like it hear in Reading H24: what sci fi are you reerring too in reading? J1: No sci fi just a question about reading H24: what about you ? H24: do you like it ? J1: Having lived here most of my life I would have to say that yes, I do like Reading H24: okay thats nice J1: Did you enjoy the rain this morning H24: yeah it was fun getting wet in the rain did yuo enjoy it ? J1: I preferred the snow H24: did it snow as well ?
Judge classification: male adult	Judge classification: machine

2.3.1 Confederate Effect

In this test J1 returned classifications of human male left, *gender blur*, and machine right, an instance of the *confederate* effect (Transcript 3). In fact they had conversed with a hidden female- male duo. Judge J1 awarded the right entity with a score of 60 out of 100 for conversational ability giving the reason: "missed some questions". The human interlocutor on the right was an international student at the time of the test. Again, cultural differences, with the male asking rather than answering questions could have swung the decision to classify them as machine.

2.4 J11: female

In the previous three transcripts the hidden pairs involved one female and one male. In the next conversation, unknown to female judge J11 they interrogated two hidden females, one was misclassified as male. J11's parallel interrogation shows a balanced conversation with both hidden interlocutors sharing the duration time almost equally (Transcript 4).

Transcript 4: J11 Interrogating two females

J11: Session 2 Round 21: simultaneously interrogating two females	
H25: female adult	H8 female
J11: Hi there H25: Hi. How are you today? J11: I'm good thanks, how are you? H25: Very well thanks. Where are you from? J11: I'm from Brighton but I live here in Reading J11: How about you? H25: I'm from Guildford. J11: Do you like it in Reading? H25: It's a nice ampus here. J11: Are you a student here? H25: No. I'm a student in Guildford. And you? J11: I was a student here but now I work here instead! J11 What do you study? H25: Sociology. You? J11: I did Psychology, and then a masters in English J11: So a similar area to you I guess H25: Ah. I'm really an economist, but I'm doing sociology now. J11: That's an interesting change, i suppose they link well together? H25: Yes. Economics is a bit narrow. Sociology takes a wider view. How did you get to chcnage? J11: I had the opportunity to do a masters for free cos I work here, and that one was in the evening!	J11: Hi there H8: hello J11 how many of these conversations have you had now? H8: 3 I think J11: Do you think anyone thinks you are a machine? H8: I hope not J11: So, where are you from? H8: Originally I'm from Swansea in WAles, but for the last few years I've been living here in Reading J11: Cool, I'm originally from Brighton but I've also been here a few years H8: Do you miss Brighton? J11: Sometimes - it's good fun and my family are there J11: Do you miss Wales? H8: Not really J11: Do you prefer badgers or squirrels? H8: Depends on the circumstance J11: What circumstances would you prefer squirrels? H8: If I was on a nice walk in the country J11: not badgers then? H8: I think they can be quite agressive J11: how so? H8: They are very protective of their homes J11: aren't you? H8: I guess so

J11: My main love was psych J11 How long have you been doing sociology? H25: So you did a masters part-time? That's hard!	J11: do you like your home? H8: It's ok
Judge classification: correct	Judge classification: male, 20s

Both hidden interlocutors posted a spelling error: "chcnage" on the left and "agressive" on the right (Transcript 4). J11 correctly identified the left hidden interlocutor as female but ranked the right hidden as "Human male British 20s". The right hidden interlocutor was in fact a private school educated female teenager. Their mature interaction could have been mistaken for masculine talk.

Post-experiment in one independent analysis of Transcript 4 by a male professor with non-first language English their view of the right interlocutor's conversation was: "*I would say that H8 is not human. ??*". In another, by a female professor with first language English, they classified the same way as female judge J11: left-female-right-male (Private emails to first author, October2015).

In the following section the reader can compare the 2human transcripts with a machine-human conversation from the same experiment.

3 MACHINE-HUMAN PAIR

Transcript 5 presents a machine-human simultaneous interrogation. In this conversation the male judge – first language English speaker, simultaneously interrogated a machine sending utterances to the left of the interrogator's screen and the human's utterances relayed to the right of the screen.

Comparing Transcripts 1-4 with Transcript 5 the uneven interaction in the latter is visible: in the machine-human test the left-side dialogue shared more content than the right-side. Hidden participant E1 utterances sent to the left of judge J13's screen show they were a loquacious interlocutor. The left entity used longer utterances than the right entity (Transcript 5). Judge J13 did not recognise this as a feature of artificial dialogue and classified the left interlocutor as human. In fact it was a computer programme, *Eugene Goostman* developed to mimic a male child of Ukrainian nationality (Demchenko and Veselov, 2008). J13 captured his rationale in a newspaper article: "I was fooled. I mistook Eugene for a real human being. In fact, and perhaps this is worse, he was so convincing that I assumed that the human being with whom I was simultaneously conversing was a computer… [Eugene] was playful, implying in his answer that he might well be a computer program whose only friends were programmers" (Pavia, 2008).

Assumption that others should *know what I know*, not knowing who Sarah Palin is (Transcript 5, right side) and perception of evasiveness as *machinelike* led J13 to classify actual human female as a machine. It is interesting that judge J13 did not ask the left interlocutor about Sarah Palin, especially as the judge uttered "You keep changing the subject" (Transcript 5-left)– a ploy of artificial dialogue systems to control and direct the conversation to their knowledge base.

Transcript 5: J13 interrogating machine-human pair

J13: Session 1 Round 9: simultaneously interrogating machine and human	
E1: Actual – Eugene Goostman machine	H20 female
J13: So let's cut straight to the point. Are you the human or the computer? E1: Some my friends are programmers. Their last project was named "Aye-Aye Engine". I still don't know what it is for - but for any input it shouts "Aye-aye, Master!", and that's all... **J13**: I think you're saying you might be either or. E1: Your opinion is very interesting... Keep going. If I'm not mistaken - you still didn't tell me where you live. OR it's a secret?:-) **J13**: You keep changing the subject. You ought to be a politician. I live in south London, a place called Brockley that no one has ever heard of. E1: South london? Hope it's a good place to live. **Interrogator**: It's very nice thanks. Where do you live? E1: I am from Ukraine, from the city called Odessa. You might have heard about it. **J13**: Yes although I don't know an awful lot about it. How did you end up here? E1: What a stupid questions you ask! Holly	**J13**: So let's cut straight to the point. Are you the human or the computer? H20: What do you think? **J13**: That's just the sort of answer I would expect from a computer. H20: Well you could be wrong!!! **J13**: That's true. It's happened a lot in the past. What do you make of Sarah Palin? H20: Sorry don't know her **J13**: How can you possibly not know her? What have you been doing for the last two months?

cow, this Loebner Prize is such a weird place! But I like it!	
Judge classification: Female adult	Judge classification: machine

4 DISCUSSION

Computer-mediated communication is an "interactive channel" allowing users to seek information in an inexpensive and efficient way (Miller 2009, in Ean, 2010). However, stereotypical views of masculinity and feminity persist, "competitiveness, assertiveness, sympathy and affection" such that "people learn sex role socialisation" (Lueptow, Garovich, Lueptow, 1995: p. 510). Dialogues in this paper show there are features distinguishing machines from human, but determining sex of a human interlocutor in short text is not clear cut. A talkative person was considered female from short text whereas another revealing cybernetics engineering study was assumed to be male (Transcript 1), possibly due to the assumption that the ratio of boys to girls taking this subject is greater.

For intelligent virtual assistants, beyond knowledge of remembering facts to maintain flowing dialogue, could more humans be engaged in education or trust in e-commerce by adding other characteristics to agents including virtual gender? One study showed "the type of character consumers preferred was most likely to be between 35-44 years old, male or female, dressed appropriately for the brand in question, animated, attractive and have a sense of humor" (Artificial Solutions: p. 3). However, in the same study younger consumers were more likely to seek older characters and "vice versa for an older audience" (p. 4). More studies are needed to examine what best suits a talking character in a robot carer looking after an elderly person in their own home.

Another issue, pointed out by De Angeli and Carpenter (2005), is that of intentionally offending a hidden interlocutor. They presented evidence of abuse found in a "corpus of spontaneous conversations" with Carpenter's online chatbot *Jabberwacky* (p. 20). This adverse factor in computer mediated communication affects humans too: despite "Teens will put up with it because technology is cool and crazy" (Bluestein, 2003 in Faulkner and Culwin, 2005). Information communication technologies enable tools "to inflict harm on women and girls" through online abuse or *trolling* (UN Broadband, 2015). Is it wiser then to develop gender-neutral agents to mitigate abuse of conversational agents? In one experiment with 24human-human control pairs, half the time the interrogators incorrectly classified male as female and vice versa. We presented four of those wrong simultaneous dialogues to shed light on why judgements were made in a particular way.

5 CONCLUSIONS

The text-talk presented in the five simultaneous Turing test dialogues in Transcripts 1-5 show the human participants revealed feelings of excitement (Transcript 1-left), disclosed personal information - judge J10 revealed they were a Reporter (Transcript 1-right), shared knowledge about places – Earley, Cardiff, Reading (Transcript 1-left), and raised awareness -badgers can be aggressive (Transcript 5-right). Gender-blur was evident in interrogator misclassifications: males hidden humans were classified female, and vice versa. Aditionally a machine programmed to imitate a male child was deemd a female (Transcript 5). Judges with first-language-English and non-first language English succumbed to gender blur. These classifications could be as a result of a) steretypical beliefs; b) disruption to expectation due to culture, or c) an unconscious bias influencing assignment of male or female characteristics to hidden interlocutors. Lastly, first impression of short text interrogation produced overall 50% correct sex classification of the human foils. Futher evaluation using statistical tools could reveal trends accompanying linguistic comparison.

6 FUTURE WORK

Analysis is ongoing of over 700 conversations realised from 426 Turing tests involving over 80 human judges, six machines and more than 50 human foils. In addition to *gender blur*, misclassifying a male as female and vice versa, the authors are evaluating male vs. female and age ranges of interrogator judges to find if there is a particular group more susceptible to deception in short text. Results will be presented in future publications.

REFERENCES

Albrechtsen, J.S., Meissner, C.A., and Susa, K.J. (2009). Can intuition improve deception detection

performance? *Journal of Experimental Social Psychology*. Vol. 45: pp. 1052-1055

Artificial Solutions. *The Top Traits of Intelligent Virtual Assistants*. White Paper. Available here: http://www.artificial-solutions.com/about-artificial-solutions/resources/registered-whitepapers/ accessed 16.10.15

De Angeli, A., and Carpenter, R., 2005. Stupid Computer! Abuse and Social Identity. Proceedings of Workshop on *Abuse: the darker side of human-computer interaction*. September 12: Rome. Available here: http://www.agentabuse.org/Abuse_Workshop_WS5.pdf

Ean, L.C., 2010. Face-to-face versus computer-mediated communication: exploring employees' preference of effective employee communication channel. *International Journal for the Advancement of Science & Arts*. Vol 1(2), 38-48

Faulkner, X., and Culwin, F., 2005. When Fingers do the Talking: a study of text messaging. *Interacting with Computers*. Vol. 17, 167-185

Holbrook, C., Fessler, D.M.T., and Navarrete, C.D., 2015. Looming large in other's eyes: Racial stereotypes illuminate dual adaptations for representing threats versus prestige as physical size. *Evolution and Human Behaviour*. In press, available from DOI: http://dx.doi.org/10.1016/j.evolhumbehav.2015.08.004

Holmes, J., 2006. Sharing a Laugh: Pragmatic aspects of humor and gender in the workplace. *Journal of Pragmatics*. Vol 38, 26-50

Holmes, J., 2005. Leadership Talk: How do leaders 'do mentoring' and is gender relevant? *Journal of Pragmatics*. Vol 37, 1779-1800

Independent, 2015a. Theresa May's speech to the Conservative Party Conference- in Full. 6 October Available here: http://www.independent.co.uk/news/uk/politics/theresa-may-s-speech-to-the-conservative-party-conference-in-full-a6681901.html

Independent, 2015b. Tory Party Conference 2015: David Cameron's speech in full. 7 October. Available here: http://www.independent.co.uk/news/uk/politics/tory-party-conference-2015-david-camerons-speech-in-full-a6684656.html

Loebner Prize, 2008. 18th Loebner Prize for Artificial Intelligence. http://loebner.net/Prizef/2008_Contest/loebner-prize-2008.html

Lueptow, L.B., Garovich, L., and Lueptow, M.B., 1995. The persistence of gender stereotypes in the face of changing sex roles: Evidence contrary to the sociocultural model. *Ethology and Sociobiology*. Vol. 16 (6), 509-530

Pavia, W., 2008. Machine Takes on Man at Mass Turing Test. *The Times UK*. Available online: http://technology.timesonline.co.uk/tol/news/tech_and_web/article4934858.ece

Reading University, 2012. Computer or Human – Can you tell the difference? https://www.reading.ac.uk/news-and-events/releases/PR451417.aspx

Shah, H., 2010. Deception detection and machine intelligence in practical Turing tests. PhD thesis, University of Reading, UK

Shah, H., Warwick, K., Bland, I.M., Chapman, C.D., and Allen, M., 2012. Turing's Imitation Game: Role of Error-making in Intelligent Thought. *Turing in Context II*, Brussels: 10-12 October

Shah, H., and Warwick, K., 2010a. From the Buzzing in Turing's Head to Machine Intelligence Contests. *Proceedings of Symposium for Towards a Comprehensive Intelligence Test.* AISB Convention, De Montfort, UK, 29 March – 1 April.

Shah, H., and Warwick, K., 2010b. Hidden Interlocutor Misidentification in Practical Turing tests. *Minds and Machines*, Vol 20(3), 441-454

Shah, H., and Warwick, K., 2008. Can Machines think? Results from the 18th Loebner Prize for Artificial Intelligence contest. The University of Reading: http://www.reading.ac.uk/15/research/ResearchReviewonline/featuresnews/res-featureloebner.aspx accessed: 7.10.15

Turing, A.M., 1952 in Braithwaite, R., Jefferson, G., Turing, A.M., and Newman, M. Can Automatic Calculating Machines Be Said to Think? Transcript of BBC radio broadcast. In (Eds) S.B. Cooper & J. van Leeuwen, *Alan Turing: His Work and Impact*, Elsevier, 2013

Turing, A.M., 1950. Computing Machinery and Intelligence. *MIND*, Vol 59 (236), pp. 433-460

UN Broadband Commission, 2015. Cyber violence against women and girls: a world-wide wake-up call. *UN Digital Development Working Group on Broadband and Gender*. Report available from: http://www.broadbandcommission.org/publications/Pages/bb-and-gender-2015.aspx accessed 7.10.15

Demchenko, E, and Veselov, V., 2008. Who Fools Whom? The Great Mystification, or Methodological Issues on Making Fools of Human Beings. In (Eds) Epstein, R., Roberts, G., and Beber, G. *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Springer

Warwick, K., and Shah, H., 2015. The importance of a human viewpoint on computer natural language capabilities: a Turing test perspective. *AI and Society* [in press]. Available from http://dx.doi.org/10.1007/s00146-015-0588-5

Warwick, K., and Shah, H., 2014. Good Machine Performance in Turing's Imitation Game. *IEEE Transactions on Computational Intelligence and AI in Games* 6 (3), 289-299

Willis, J., and Todorov, A. (2006). First Impressions: Making up your mind after 100ms exposure to a face. *Psychological Science* 17 (7), pp 592-598