Information Design Journal

Johan F. Hoorn and Matthijs A. Pontier

Robot communication - human contact with androids

Five challenges face the design of lifelike humanoid robots – also called androids. Robots should live up to social needs – people are getting older, lonelier, and look for alternatives to face-to-face contact, for instance, in coaching and therapy (1). Surrogate partners substantiate the imaginative friends that people silently cherish and robot designers assume that the better they can simulate social and affective behavior, the more effective the robotic partner is. Yet, what should be regarded as natural behavior and does maximal human-likeness equal optimal user satisfaction (2)? In addition, multidisciplinary design teams prompt transdisciplinary theory to create mutual understanding and cover the wide-ranging facets of robot design. Yet, how to combine the diverse theories from various domains when confronted with major unification problems (3)? Empirical approaches to robot design usually miss out on logic consistency of the theory. As a mirror image, formalization and mathematical modeling often lack external validity. Verification before validation is the new way to go but not necessarily the easiest one (4). Finally, if we do succeed at creating lifelike, affective, sociable androids, the cognitive models underlying their behavior could be turned into spyware and malware (5). Therefore, this paper ends on an ethical note.

Societal need for android contact

That society becomes older and individuals more self-centred are two trends that contribute to the experience that emotionally and socially, people become more isolated (Killeen, 1998). In reaction, scientists develop conversational agents that act as friends for geriatric patients – as a kind of artificial parrot. The robot generates fiction stories in which the characters show realistic behavior and the user can influence the storyline. Virtual storytelling has become a hot topic in academic areas such as artificial intelligence, computer-human interaction, and cognitive science but the industry shows an interest as well with respect to game design and screen writing (Romano et al., 2007).

Androids telling stories are not only useful to make the elderly feel less lonely. Virtual storytelling proved to be useful for clinical experts in the treatment of behavior problems, family counseling, and training (e.g., Painter, Cook, & Silverman, 1999), acquisition of language and other sign systems (Schlosser & Lloyd, 1993), or in persuasive contexts (e.g., Lee & Leets, 2002). The motivation for developing robots that tell virtual stories may vary from pure entertainment (e.g., computer games) to educational purposes (e.g., training environments, teaching strategies) (Johnson et al., 2005), persuasion (e.g., science and health communication), or clinical therapy.

With regard to mental health, many self-help programs are delivered through the Internet (Christensen et al., 2004; Clarke et al., 2002; Spek et al., 2007). Several studies concluded that self-help therapies are useful and efficient in reducing mental health problems (e.g., Anderson et al., 2005; Adopaca & Miller, 2003; Den Boer et al., 2004; Spek et al., 2007). Compared to traditional therapy methods, Web-based self-help may even be more efficient and less expensive (Andrews et al., 2001; Bijl & Ravelli, 2000; Griffiths et al., 2006). Self-help therapy may be even more successful when the interface is enhanced or replaced by a robot therapist. The anonymity of robot-supported self-help therapy could overcome potential embarrassment of undergoing formal treatment (cf. Peck, 2007; Rochlen, Zack, & Speyer, 2004; Williams, 2001).

Android contact can accommodate those who are lonely, provide health advice to the elderly, makes games more interesting and online instruction livelier, is useful for coaching, counseling, and self-help therapy. In extreme circumstances, robots can be the better self of human operators in executing dangerous tasks. For example, NASA initiated the P2P-HRI program to explore the possibility of robots and humans collaborating as teammates, while perceiving each other as peers (Fong & Nourbakhsh, 2004). When regular therapy puts up too high a threshold, a robot therapist is less threatening, what the patient reveals is inconsequential, the patient is in control, and all in all, interaction with the robot therapist has a "dear diary" effect.

Particularly in self-help therapy, robot designers face the challenge to make an android that can understand and simulate genuine affective behavior. Ironically and contrary to what designers usually attempt, the fact that the robot is not exactly human and thus, is not completely natural, makes people probably less shy and confide their secrets to their unnatural conversation partner. It seems then that imperfect representations of reality may have its benefits but…

How uncanny can the valley be?

What is natural behavior? People are happier and live longer by taking care of a pet or a plant. They smile at their pencil, pet their car on the back, and give their dish washer a name (cf. Brave and Nass, 2002; Reeves & Nass, 1996). Technological animism provides great opportunities for robots to raise sympathy without having to be completely humanlike. Yet, the first conversational agents and social robots (e.g., ELIZA: Weizenbaum, 1966) were seen as a major technical breakthrough but not as successful conversation partners. Later, more sociable robots such as Kismet featured smiles (supposedly expressing happiness) and frowns (supposedly expressing sadness) (Breazeal, 2002).

At the China Robot Expo in Beijing in October 2006, I ran into Zou Ren Ti, founder of the Xi'an Superman Sculpture Research Council, or actually, I bumped into Zou Ren Ti, the robotic look-alike of its inventor. I was startled. Zou the robot talked, moved, and had a skin of silica gel that gave him this realistic look-and-feel (Hamilton et al., 2006).

Using the framework of the "uncanny valley" (Mori, 2005/1970), ELIZA was still close to an industrial robot: not too humanlike and therefore making a rather unfamiliar impression. Sociable robots such as Kismet are humanlike enough to reach a peak in experienced familiarity. But then the uncanny valley lies ahead. If you make a robot too humanlike but not yet perfect, so Mori argued, it comes close to the experience of dead people and users will resign from it.

So where is the optimum? How natural should a robot behave emotionally so that it is acceptable to a human without becoming so realistic that it is threatening again? Zou Ren Ti frightened me. If we want to design humanlike conversation partners and create a robot that behaves emotionally in a natural way, the "dear diary" effect for self-help therapy may be gone.

In our own research (Bosse et al., 2008; Hoorn, 2008), we too try to mimic human affective behavior as closely as possible. To create complex personalities with humanlike properties, virtual characters should be capable of mixed responses to their human user ("Although you try your best, you treat me badly"). In our research, embodied conversational agents are equipped with an integration of the emotion regulation model by Gross (1998; 2001) and the I-PEFiC model of user engagement with virtual characters by Van Vugt et al. (2006; 2007; 2008). We model approach and avoidance tendencies and translate them to the robot domain. Increasingly, the software can recognize and mimic the experiential ambiguities that are typical for human affect (so called involvement-distance trade-offs), which transpire because the robot considers a user to be beautiful but at the same time unskilled. The affective structure that our robots build up feeds into the formal models we distilled from theories of emotion regulation so that the robot can determine to continue the conversation, walk away, or change the situation.

Transdisciplinary theory – misapprehension and non-acceptance

Design of an android is not a technical matter alone. Apart form artificial intelligence and software engineering for multimedia hardware, theories and models of human life are as important to explain communication rules, social interaction and perception, or the appraisal of certain social situations. In mediated interpersonal communication (Konijn & Van Vugt, 2008) and media psychology, emotions play a salient role and cover an important area of research. Other relevant disciplines are emotion psychology, computer-mediated communication, and computer-human interaction. The boundaries that are crossed are not only within but also across schools and this brings about misapprehension and non-acceptance of the scientific tradition that the researcher is unfamiliar with.

The combination of diverse theories from various disciplines inevitably forces to reconsider the theoretical framework and methods one is used to. Because there is a problem in unifying the different points of view and the danger of drowning in terminological confusion, a simple integration or adaptation of theories is insufficient. Theory that overarches the sub disciplines is germane but hard to get accepted because it seems to downplay the importance of well-established traditions. Authorities suddenly have to admit that they oversee but a part of the bigger picture.

To build emotion regulation into robots, we formalized the informal model by Gross (1998) as our point of departure. This model describes five types of strategies humans use to adapt their emotional response levels: (1) situation selection, (2) situation modification, (3) attentional deployment, (4) cognitive change, and (5) response modulation. Because of the high abstraction level of formal languages, we were capable of linking emotion regulation with the involvement-distance trade-offs that the user-engagement model (I-PEFiC) describes (Van Vugt et al., 2006; 2007; Van Vugt, 2008). Whereas the model on user engagement focuses on the *elicitation* of emotion, the affect regulation model addresses the *regulation* of emotion. On a conceptual level, we could bring the two theories together, because the satisfaction with the user that is generated as output of the involvement-distance trade-off can almost directly be used as input for affective situation selection. Current satisfaction is checked for a certain threshold and if it is too low, the robot will evaluate its expected satisfaction in alternative situations (e.g., choose another user).

Verification before validation

Theories of communication and psychology are often well-founded on empirical data but often lack internal consistency. As a counterpart, formal languages and logic models are internally consistent but often lack ecological validity. Ideally, informal models that are based on empirical data are formalized mathematically so that all parameters are explicated. Simulation tests verify the internal consistency of the theory after which experiments with real people validate what the theory predicts. Doing it both ways is unusual in the social sciences and in computer science but should be common practice – at least in robot development.

Computational models are able to imitate the development of cognitive processes over time. This process is called simulation, and models that can be used to simulate behavior over time are called dynamic models. The result of a simulation is a sequence of states of the model at subsequent time points, which is called a trace. Simulation experiments with these models can be used to improve the understanding of the processes and to make predictions. The behavior of the processes is studied by looking at a number of simulation traces under different circumstances. The modeling and simulation cycle consists of four activities:

1. Conceptualization: determining the main aspects and their relations
2. Formalization: specifying the detailed model
3. Simulation: performing experiments with the model to generate traces
4. Evaluation: verifying whether the model behaves as expected and validating the applicability of the model in practice

If the evaluation reveals that the model is not correct, the process starts again with a new conceptualization phase. When the simulation experiments show that the robot generates realistic communicative behavior, the conversational agent is visualized in a graphical environment or built in 3D (cf. Zou Ren Ti).

In a ‘Turing Test’ (Turing, 1964) experiment, participants play, for example, a digital game and judge the human quality of the behavior of their opponent or teammate. In some cases, the virtual other will be another human player, whereas in other cases, the virtual other will be a robot equipped with our cognitive models. If the participants cannot distinguish between a human and a robot player, the robot has passed the Turing Test. If the robot failed the test, redesign is needed. In a complementary experiment, the Wizard of Oz technique (Landauer, 1987) ensures that certain robot functionality such as natural language skills is performed by a human to make the robot look more autonomous and humanlike. If then again users cannot detect differences in the quality of behavior with or without the human helper, the robot passed its android exam.

Robots in disguise – ethical considerations

Androids are robots that act and look like humans. What if we make it through the uncanny valley (Mori, 1970/2005) and make an electronic version of a healthy person? Turing Test experiments will show that the android is not distinguishable from humans on a range of communication variables (e.g., trust, persuasion, credibility). The Wizard of Oz technique will indicate that the human who mimics the robot is taken for an android. So now that we have it – the robot cannot be distinguished any more – are we in good shape?

Androids that communicate affectively with the user by showing emotions have the potential to mislead people, if programmed that way. Androids could lie while perfectly showing fake emotions. If robots can adapt to the user's needs, software could be developed that has the purpose to frustrate the user. Security is also an issue. When androids have the opportunity to monitor their user affectively, unauthorized people could be interested in this information as well. The information the robot keeps about its users should be stored in a secure way, so it is inaccessible to third parties. When we apply this to self-help therapy, android therapists will keep the Medical Ethics Review Committee busy for a long time.

Conclusions

Contact with androids can support people's wellbeing, for example, as a pet, a toy, as teacher, storyteller, coach, or therapist. That androids are not identical to humans helps people to confide in them because the information people disclose about themselves is "safe." This leads to several considerations. How realistic should a robot be? Are people willing to talk to a tin can or do they prefer an exact copy of a human being? People may think that a non-human counsellor is safe but it is just another piece of software so that information can be used for good or for ill. Owing to its multidisciplinary nature, robot science and design have to overcome many communication barriers – in theory and jargon – and in accepting the importance of unfamiliar points of view. One such example is that communication and psychology should learn how to model their theories mathematically whereas computer science should learn how to test their mathematical models in real life. Only then will robots emerge that can be truly called androids.

References

Anderson, L., Lewis, G., Araya, R., Elgie, R., Harrison, G., Proudfoot, J., Schmidt, U., Sharp, D., Weightman, A., & Williams, C. (2005). Self-help books for depression: how can practitioners and patients make the right choice? *British Journal of General Practice, 55, 514,* 387-392.

Andrews, G., Henderson, S., & Hall, W. (2001). Prevalence, comorbidity, disability and service utilisation. Overview of the Australian National Mental Health Survey. *The British Journal of Psychiatry, 178, 2,* 145-153.

Bijl, R.V., & Ravelli, A. (2000). Psychiatric morbidity, service use, and need for care in the general population: results of The Netherlands Mental Health Survey and Incidence Study. *American Journal of Public Health, 90, 4,* 602-607.

Bosse, T., Hoorn, J.F., Pontier, M.A., & Siddiqui, G.F. (2008). A robot's experience of another robot: Simulation. In V. Sloutsky, B.C. Love, & K. McRae (Eds.), *Proceedings of the 30th International Annual Conference of the Cognitive Science Society, CogSci'08* (pp. #-#). Washington DC.

Brave, S., & Nass, C. (2002). Emotion in human-computer interaction. In J. Jacko & A. Sears (Eds.), *Handbook of Human-Computer Interaction* (pp. 251-271). Mahwah, NJ: Erlbaum.

Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: MIT.

Christensen, H., Griffiths, K.M., & Jorm, A.F. (2004). Delivering interventions for depression by using the internet: randomised controlled trial. *BMJ, 328,* 265-269.

Clarke, G., Reid, E., Eubanks, D., O'Connor, E., DeBar, L.L., Kelleher, C., Lynch, F., & Nunley, S. (2002). Overcoming depression on the Internet (ODIN): a randomized controlled trial of an Internet depression skills intervention program. *Journal of Medical Internet Research, 4, 3,* e14.

Den Boer, P.C.A.M., Wiersma, D., & Van den Bosch, R.J. (2004). Why is self-help neglected in the treatment of emotional disorders? A meta-analysis. *Psychological Medicine, 34, 6,* 959-971.

Fong, T., & Nourbakhsh, I. (2004). Peer-to-peer human-robot interaction for space exploration. *AAAI Fall Symposium*.

Griffiths, F., Lindenmeyer, A., Powell, J., Lowe, P., & Thorogood, M. (2006). Why are health care interventions delivered over the Internet? A systematic review of the published literature. *Journal of Medical Internet Research, 8, 2,* e10.

Gross, J.J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2, 3,* 271-299.

Gross, J.J. (2001). Emotion regulation in adulthood: Timing is everything. *Current Directions in Psychological Science, 10, 6,* 214-219.

Hamilton, A., Murray Buechner, M., Caplan, J., Rothman, W., & Park, A. (2006). Best inventions 2006. *Time*. Available at http://www.time.com/time/2006/techguide/bestinventions/inventions/toys5.html

Hoorn, J.F. (2008). A robot's experience of its user: Theory. In V. Sloutsky, B.C. Love, & K. McRae (Eds.), *Proceedings of the 30th International Annual Conference of the Cognitive Science Society, CogSci'08* (pp. #-#). Washington DC.

Johnson, W.L., Vilhjalmsson, H., & Marsella, S., (2005). Serious games for language learning: How much game, how much AI? In C-K. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology* (pp. 306-313). Amsterdam: IOS.

Killeen, C. (1998). Loneliness: an epidemic in modern society. *Journal of Advanced Nursing, 28, 4,* 762-770.

Konijn, E.A., & Van Vugt, H.C. (2008). Emotions in mediated interpersonal communication. Toward modeling emotion in virtual humans. In E.A. Konijn, S. Utz, M. Tanis, & S.B. Barnes, (Eds.), *Mediated Interpersonal Communication.* New York: Taylor & Francis Group / Routledge.

Landauer, T.K., (1987). Psychology as a mother of invention. *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface*, pp. 333-335.

Lee, E. & Leets, L. (2002). Persuasive storytelling by hate groups online: examining its effects on adolescents. *American Behavioral Scientist, 45,* 927-957.

Mori, M. (2005). Bukimi no tani [The uncanny valley] (K.F. MacDorman & T. Minato, Trans.). Retrieved from http://www.theuncannyvalley.com (Originally published in 1970; *Energy, 7, 4,* 33-35).

Painter, L.T., Cook, J. W., & Silverman, P.S. (1999). The effects of therapeutic storytelling and behavioral parent training on noncompliant behavior in young boys. *Child and Family Behavior Therapy, 21, 2,* 47-66.

Peck, D. (2007). Computer-guided cognitive–behavioural therapy for anxiety states. *Emerging Areas in Anxiety, 6, 4,* 166-169.

Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York: Cambridge University.

Rochlen, A.B., Zack, J.S., & Speyer, C. (2004). Online therapy: Review of relevant definitions, debates, and current empirical support. *Journal of Clinical Psychology, 60, 3,* 269-283.

Romano, D.M., Brna, P., Robertson, J., & Louchart, S. (2007). AI and narrative games for education. In P. Olivier & C. Kray (Eds.), *Artificial & Ambient Intelligence. Proceedings of the AISB Annual Convention, AISB'07*. Newcastle upon Tyne, pp. 385.

Schlosser, R.W., and Lloyd, L.L., (1993). Effects of initial element teaching in a story-telling context on Blissymbol acquisition and generalization. *Journal of Speech and Hearing Research, 36,* 979-995.

Turing, A.M. (1964). Computing machinery and intelligence. In A. Anderson (Ed.), *Minds and Machines*. Engelwood Cliffs, NJ: Prentice Hall.

Van Vugt, H.C. (2008). *Embodied agents from a user's perspective*. Doctoral dissertation, VU University, Amsterdam.

Van Vugt, H.C., Hoorn, J.F., Konijn, E.A., & De Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies, 64, 9,* 874-888. DOI: 10.1016/j.ijhcs.2006.04.008

Van Vugt, H.C., Konijn, E.A., Hoorn, J.F., Eliëns, A., & Keur, I. (2007). Realism is not all! User engagement with task-related interface characters. *Interacting with Computers, 19, 2,* 267-280.

Weizenbaum, J. (1966). ELIZA - A computer program for the study of natural language communications between men and machines. *Communications of the Association for Computing Machinery, 9,* 36-45.

Williams, C. (2001). Use of written cognitive-behavioural therapy self-help materials to treat depression. *Advances in Psychiatric Treatment, 7,* 233-240.

Contact

Johan F. Hoorn
Center for Advanced Media Research Amsterdam (CAMeRA)
VU University
De Boelelaan 1081
Amsterdam, 1081 HV
Email: jf.hoorn@camera.vu.nl