Establishing Human Personality Metrics for Adaptable Robots During Learning Tasks

Cory J. Hayes and Laurel D. Riek
Department of Computer Science and Engineering
University of Notre Dame, Notre Dame, IN, USA

Abstract

This paper describes our ongoing research effort to explore how personality types factor into HRI; in particular, the degree of patience a person has when teaching an error-prone robot in a learning from demonstration setting. Our goal is to establish personality metrics that will ultimately allow for the design of algorithms that automatically tune robot behavior to best suit user preferences based on personality.

Introduction

Robots are expected to become a ubiquitous technology that will have significant impacts on society (Computing Community Consortium 2013). As such, robots will be situated in a large array of domains such as personal assistance, service, education, and leisure. Robots will be expected to perform a variety of tasks within these domains, such as household chores and medical assistance (Ray, Mondada, and Siegwart 2008). While there may be a considerable amount of time before this future society becomes a reality, it is important for HRI researchers to consider the implications of this scenario and alternate situations.

In particular, as robotics technology proliferates it is likely that robots will not be universally accepted. There are several reasons for this, such as robot appearance, perceived usefulness, and perceived ease of use. The notion that a robot's appearance can influence how people perceive and respond to it has been well-established in HRI literature (Goetz, Kiesler, and Powers 2003). Perceived usefulness and perceived ease of use serve as the main components of the Technology Acceptance Model (TAM) (Davis 1989), which models the tendency of people to use a certain technology, and this phenomenon may be compounded by the overall lack of interactions that the general public has had with robots (Riek, Adams, and Robinson 2011).

Another reason robotics technology may not be universally accepted is due to individual differences, which is the focus of our work. People have different backgrounds that can uniquely affect their attitudes towards technology. Some individuals may come from cultures where robots are treated as living entities with respect, while others may come from cultures that view robots more as tools (Kitano 2005). People also have a wide range of cognitive and physical abilities that can affect how they perceive, interact with, and accept robots. Additional attributes that may factor into a person's degree of robot acceptance include age, gender, educational level, perceptions of job security, media exposure to robots, and previous encounters with robots.

Specifically, we focus on the individual differences related to personality. Personality can affect how people interact with different technologies, as seen in both the HCI and HRI literature. In HRI, personality has been explored prominently in the domain of anthropomorphism (Fussell et al. 2008; Fong, Nourbakhsh, and Dautenhahn 2003; Dautenhahn 2004). HRI researchers have also focused on user responses to robots who exhibit some form of personality (Lee et al. 2006), and have found that users respond more positively towards robots that exhibit personality qualities similar to theirs (Aly and Tapus 2013). Personality has also been shown to affect the willingness of people to accept a new technology as well as their willingness to adapt to it (Davis 1989; Devaraj, Easley, and Crant 2008).

Robots will be expected to perform a wide range of tasks that vary in complexity, and they will undoubtedly make mistakes. To help facilitate acceptance of robots, it is important to anticipate how people will respond when a robot makes a mistake. However, it is not feasible for robot designers to explicitly account for every single type of person a robot will encounter. Lee and colleagues (Lee et al. 2010) suggest that robot mistakes can be expected and mitigated, and some in the robotics community argue this can be accomplished via end-user programming, e.g., using Learning from Demonstration (LfD) (Argall et al. 2009).

In LfD, a robot learns to create a mapping (policy) between specific actions and world states from watching a teacher perform these actions in demonstrations. By following the actions of the teacher, the robot automatically learns to reproduce these actions. The main benefit of LfD is that it does not require the teacher to have specialized skills (Konidaris et al. 2011), which is beneficial for members of the general population with limited or no experience in programming. In this approach, humans can teach and correct robots, tailoring the robot's behavior to their expectations. However, LfD approaches assume people are willing to patiently teach robots, which may not be the case.

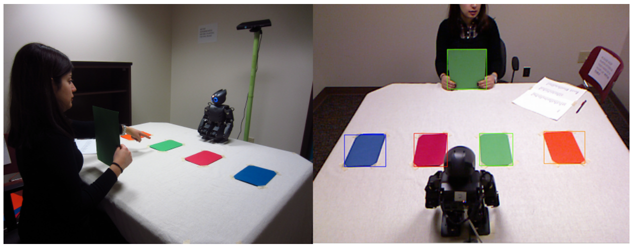

Figure 1: In one study we conducted, a participant teaches an autonomous robot to identify four colors.

Thus, in our work, this motivates several research questions regarding the relationship between an individual's personality traits and responses to robot mistakes. In particular, we are interested to know how personality traits affect a person's patience when dealing with robot errors. We also wonder how these traits affect user satisfaction and, to a greater extent, robot acceptance.

Experimental Paradigm

We have designed an experimental paradigm using LfD to explore the above research questions. In our LfD task, participants teach an autonomous humanoid robot (DARwIn-OP) to identify four colors using large color cards. Each experiment session consists of one participant directly interacting with the robot alone in a room (see Figure 1). We chose a color identification task since it was fail-safe approach from an autonomy perspective, given the controlled room conditions.

Participants take a personality test prior to undergoing the LfD task. In particular, we employ measures that utilize the five-factor model of personality. This model generalizes personality into the five dimensions of openness, conscientiousness, extraversion, agreeableness, and neuroticism (Digman 1990), and each dimension has an associated scale.

The LfD task occurs in two phases: learning and testing. In the learning phase, the participant teaches the robot to verbally identify and point to a color. The participant teaches the robot each color using the following sequence: 1) pick up the card of a certain color, 2) hold it in front of the robot, 3) state the color of the card, 4) point to the card of the same color on the table, and 5) wait for the robot to acknowledge that it learned the color by it pointing to the same card on the table and verbally stating the color.

After the learning phase, the participant then tests the robot's ability to identify colors. The testing phase consists of the participant holding up a color card and waiting for the robot to point to and state the correct color card. The goal of this phase is for the robot to correctly identify a consecutive number of colors within a time limit.

We inform the participants beforehand that since the robot is basing its knowledge from limited training, it may occasionally make a mistake during identification. If this error does occur, the participant is instructed to correct the robot by once again holding up the correct color and stating it. Though participants are led to believe that they are teaching the robot, in truth all of the robot's behaviors, including errors, are pre-programmed. This autonomous behavior is accomplished through a blob detection program using the ROS *cmvision* package and RGB data from a Microsoft Kinect, which transmits commands to the robot.

The testing phase is the stage where we intentionally make manipulations to the experiment to explore how personality affects interaction with the robot. In this phase, the robot is explicitly programmed to make numerous mistakes. As an example, one mistake the robot could make is correctly stating that the color card held by the participant is green while it points to the blue color card on the table. This phase is the key part of the experiment to determine if there are correlations between specific personality types and certain observable actions by the participants.

Future Work

We plan to conduct a series of experiments utilizing this paradigm. For the first set of experiments, we focus on how personality traits affect patience when dealing with an error-prone robot. Patience can be measured in several ways. Initially, we are exploring the degree to which participants are willing to correct the robot. Additional manipulations to the experiment may be made; for example, we may vary the types of errors the robot makes, such as the degree of induced frustration (Klein, Moon, and Picard 2002) and the types of mitigation strategies employed (Lee et al. 2010).

The relationships between observable behavior and specific personality dimensions will help establish personality metrics in HRI. Establishing these metrics are a necessary precursor to developing adaptable systems in robots that can automatically detect user personality through interaction and modify their behavior accordingly.

Though we are initially focusing on human patience, additional qualities such as trust, may be explored. Trust is an emerging area of focus in HRI, and one of the factors that affects a person's trust in a robot is reliability. If people perceive a robot to be unreliable, they may opt to not use the robot or give it less autonomy (Desai et al. 2009). The robot in our current experimental setup is programmed to make numerous errors and may be perceived to be less reliable and, to a greater extent, less trustworthy. However, even if a robot makes errors, trust can be sustained by having the robot display varying degrees of confidence in its decisions (Desai et al. 2013), so the manipulations to our experimental setup with this focus may vary slightly from our current ones.

The overarching goal of this line of research is to allow for the creation of adaptable robot behavioral systems that can take a user's personality into account. If successful, robots would be able to dynamically learn users' specific personality profiles through repeated interactions, and fine tune their behavior accordingly. This personality module could be one of many modules, each with their own specialized focus, that work in conjunction to create a robust robot behavioral system, ultimately leading to the creation of truly personalized social robots.

References

Aly, A., and Tapus, A. 2013. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* 325–332.

Argall, B. D.; Chernova, S.; Veloso, M.; and Browning, B. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57(5):469–483.

Computing Community Consortium. 2013. A Roadmap for US Robotics From Internet to Robotics. Technical report.

Dautenhahn, K. 2004. Robots we like to live with?! - a developmental perspective on a personalized, life-long robot companion. *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication* 17–22.

Davis, F. D. 1989. Perceived Usefulness, Perceived Ease Of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13(3):319–340.

Desai, M.; Stubbs, K.; Steinfeld, A.; and Yanco, H. 2009. Creating Trustworthy Robots: Lessons and Inspirations from Automated Systems. *Proceedings of the AISB Convention: New Fronteiers in Human-Robot Interaction*.

Desai, M.; Kaniarasu, P.; Medvedev, M.; Steinfeld, A.; and Yanco, H. 2013. Impact of robot failures and feedback on real-time trust. *8th ACM/IEEE International Conference on Human-Robot Interaction* 251–258.

Devaraj, S.; Easley, R. F.; and Crant, J. M. 2008. Research Note—How Does Personality Matter? Relating the Five-Factor Model to Technology Acceptance and Use. *Information Systems Research* 19(1):93–105.

Digman, J. M. 1990. Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology* 41(1):417–440.

Fong, T.; Nourbakhsh, I.; and Dautenhahn, K. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42(3-4):143–166.

Fussell, S. R.; Kiesler, S.; Setlock, L. D.; and Yew, V. 2008. How people anthropomorphize robots. *Proceedings of the 3rd international conference on Human robot interaction* 145–152.

Goetz, J.; Kiesler, S.; and Powers, A. 2003. Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation. *12th IEEE International Workshop on Robot and Human Interactive Communication* 55–60.

Kitano, N. 2005. Roboethics - a comparative analysis of social acceptance of robots between the West and Japan. *The Waseda Journal of Social Sciences* 6.

Klein, J.; Moon, Y.; and Picard, R. W. 2002. This computer responds to user frustration: Theory, design, and results. *Interacting with computers* 14.

Konidaris, G.; Kuindersma, S.; Grupen, R.; and Barto, A. 2011. Robot learning from demonstration by constructing skill trees. *The International Journal of Robotics Research* 31(3):360–375.

Lee, K. M.; Peng, W.; Jin, S.-A.; and Yan, C. 2006. Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human-Robot Interaction. *Journal of Communication* 56(4):754–772.

Lee, M. K.; Kiesler, S.; Forlizzi, J.; Srinivasa, S.; and Rybski, P. 2010. Gracefully mitigating breakdowns in robotic services. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* 203–210.

Ray, C.; Mondada, F.; and Siegwart, R. 2008. What do people expect from robots? *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems* 3816–3821.

Riek, L.; Adams, A.; and Robinson, P. 2011. Exposure to Cinematic Depictions of Robots and Attitudes Towards Them. *ACM/IEEE Conference on Human-Robot Interaction (HRI Workshop on Expectations and Intuitive Human-Robot Interaction).*