Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research xxx (2016) xxx–xxx

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

The effects of cognitive biases and imperfectness in long-term robot-human interactions: Case studies using five cognitive biases on three robots

M. Biswas *, J. Murray

University of Lincoln, School of Computer Science, Lincoln, UK

Received 14 April 2016; accepted 31 July 2016

Abstract

The research presented in this paper demonstrates a model for aiding human-robot companionship based on the principle of 'human' cognitive biases applied to a robot. The aim of this work was to study how cognitive biases can affect human-robot companionship in long-time. In the current paper, we show comparative results of the experiments using five biased algorithms in three different robots such as ERWIN, MyKeepon and MARC. The results were analysed to determine what difference if any of biased vs unbiased interaction has on the interaction with the robot and if the participants were able to form any kind of 'preference' towards the different algorithms. The experimental presented show that the participants have more of a preference towards the biased algorithm interactions than the robot without the bias.

Keywords: Human-robot interaction; Human-robot long-term interactions; Humanoid robot; Cognitive bias; Imperfect robots

1. Introduction

It is evident that social behaviour is an important factor in human-human, and then we can be safe to assume that such interactions are important in social cognition behaviours in social robots during robot-human interactions. Mahani and Eklundh (2009) suggest that, "If through long-term use these [service] robots gain social skills, they could be supportive of some social roles that people might assign to them". To develop such social intelligence, researchers have studied various methods for robots to adapt to human-like behaviour based social roles. Few of the most popular methods suggest developing human-like attributes in robots, such as trait based personality attributes, gesture and emotions expressions and anthropomorphism.

Walters, Syrdal, Dautenhahn, Boekhorst, and Koay (2008) investigated the identifying links between human personality and attributed robot personality where the team investigated human and robot personality traits as part of a human-robot interaction trial. Research suggests that developing cognitive personality trait attributes in robots can make them more acceptable to humans (Lee, Seung-A Jin, & Yan, 2006). In addition to this, expressing emotions and mood changes in interactions can help to make the attachment bond stronger between a human and the robot. Meerbeek, Saerbeck, and Bartneck (2009) designed an interactive personality process in robots which was based on Duffy's (2003) anthropomorphism idea. Indeed, Duffy suggests that anthropomorphic or lifelike features should be carefully designed and should be aimed

* Corresponding author.
E-mail address: mrbiswas@lincoln.ac.uk (M. Biswas).

http://dx.doi.org/10.1016/j.cogsys.2016.07.007
1389-0417/

at making the interaction with the robot more intuitive, pleasant and easy.

Reeves and Nass (2000) have shown that users will demonstrate certain biased driven personality traits to machines (e.g. Computers) and from that research they propose a 'user driven' mental model for domestic robots. Walters et al. (2008) investigated people's perceptions of different robot appearances and associated attention-seeking features in video-based Human-Robot interaction trials. Their study revealed participant's preferences for various features of the robot's appearance and behaviour with their personality attributions towards the robots being comparatively similar to their own personalities. The above studies demonstrate approaches to making a robot more humanlike and thereby more intuitive for people to interact with. It is important to consider that humans have for millennia, interacted with other humans and as such our interactions and social norms are reflective of our own personalities and behaviours. It is therefore only natural that if we wish for humans to engage and interact with robots, that these robots not only understand human social constructs, but also display these traits. The research presented in this paper investigates an approach to developing socially interactive robots by applying selected cognitive biases with the aim to providing a more humanlike interaction.

Cognitive biases play a large part in influencing a human's characteristics and behaviours (Wilke & Mata, 2012). Human personalities are considered unique but based on a set of different social behaviours, social norms and cultures (Haselton, Nettle, & Andrews, 2005). Kahneman and Tversky (1972) suggest that human thinking can be affected by a variety of biases which can influence a human into making wrong decisions, bad judgments and other fallible actions, after all we're only human!

Such differences in cognitive imperfectness among individuals hugely affect that individual's interactions, making them unique, natural and human-like. Making faults and misjudgments are common human characteristics. But in developing humanlike robots, we sometime ignore such facts and attempt to make robots as faultless as possible, with perfect memory recall and repeatable actions, that is, we make them less humanlike. Such cognitive imperfections (e.g. forgetfulness, making mistakes) have has yet not been fully explored in social robots for the purpose of developing a human-robot companionship. In the current research described in this paper we approach to find out the influences of cognitive biases in human-robot interactions by developing five cognitive biases (misattribution, empathy gap, Dunning-Kruger effects, self-serving and humorous effects) in three different robots (ERWIN, MyKeepon and MARC see Figs. 2, 8 and 13). The biases were developed individually and, based on the main attributes of such biases. To compare the biased interactions there were non-biased interactions developed as well which were made free from the selected bias effects.

2. The project: cognitive bias in human-robot interaction

Cognitive biases are often a result of an attempt to simplify information processing which can help to make sense of the world and reach decisions with relative speed (Bless, Fiedler, & Strack, 2004). Sometimes, these biases lead to poor decisions and bad judgments, but in other situations, those judgemental choices can be useful. Biases refer to a systematic pattern of deviation from rationality in judgement, whereby inference about other people and situations might be drawn in illogical fashion (Haselton et al., 2005). In a given situation however, biases can sometimes lead to a more effective set of actions (Gigerenzer & Goldstein, 1996). For example, if the given context demands immediate action over accuracy, heuristic biases enable the taking of decisions faster (Tversky & Kahneman, 1974). Cognitive biases can arise from various processes that are sometimes difficult to distinguish, such as social influence (Wang, 2001), information processing shortcuts, mental noises (Hilbert, 2012), limited brain capacity of information processing (Marios, 2005; Simon, 1955) and emotional and moral motivation (Pfister & Böhm, 2008).

Bless et al. (2004) suggested that cognitive biases can influence a human's behaviour towards positive or negative ways. Biases can affect individual's decision making (Tversky & Kahneman, 1974), behaviours (Brand, 1985/1986) and social beliefs (Huijbregts, Warren, de Sonneville, & Swaab-Barneveld, 2007). It is understood that such cognitive biases among other factors (e.g. mood, emotions, traits) affect on the individual's differences in characteristics behaviours. Society is an example of each person being different in behaviour and each has got their very own unique characteristics. In our understanding, such differences in cognitive characteristics among individuals are what make human interactions unique, natural and human-like. In existing social robotics, robots are now able to imitate different human behaviours, for example, eye-gazing, making gestures while talking, expressing emotions and others. But in human-human interactions, individual's own characteristics biases (e.g. forgetfulness, empathic gap, self-serving, humorous effects) are present which are absent in the current social robots.

Sometimes a robot's social behaviours lack that of a human's common characteristics such as idiocracy, humour and common mistakes. Many robots are able to present social behaviours in human-robot interactions but unable to show such human-like cognitively biased behaviours (e.g. forgetfulness, unable to understand correct emotions, bragging, blaming, remembering humorous events). Recent studies have focused humanlike faulty behaviours to develop in robot to find out their effects in human-robot interaction. Salem, Lakatos, Amirabdollahian, and Dautenhahn (2015) studied on how the perception of erroneous robot behaviour influences human interaction choices and the willingness to cooperate with the robot. Robinette, Li, Allen, Howard, and Wagner (2016) studied faulty behaviours in robots and 'over trust' of participants

Please cite this article in press as: Biswas, M., & Murray, J. The effects of cognitive biases and imperfectness in long-term robot-human interactions: Case studies using five cognitive biases on three robots. *Cognitive Systems Research* (2016), http://dx.doi.org/10.1016/j.