

# Modeling the Dynamics of Nonverbal Behavior on Interpersonal Trust for Human-Robot Interactions

Jin Joo Lee and Brad Knox and Cynthia Breazeal

MIT Media Lab

20 Ames st. E15-468

Cambridge, MA 02142

{jinjoo, bradknox}@mit.edu, cynthiab@media.mit.edu

## Abstract

We describe research towards creating a computational model for recognizing interpersonal trust in social interactions. We found that four negative gestural cues—leaning-backward, face-touching, hand-touching, and crossing-arms—are together predictive of lower levels of trust. Three positive gestural cues—leaning-forward, having arms-in-lap, and open-arms—are predictive of higher levels of trust. We train a probabilistic graphical model using natural social interaction data, a “Trust Hidden Markov Model” that incorporates the occurrence of these seven important gestures throughout the social interaction. This Trust HMM predicts with 69.44% accuracy whether an individual is willing to behave cooperatively or uncooperatively with their novel partner; in comparison, a gesture-ignorant model achieves 63.89% accuracy. We attempt to automate this recognition process by detecting those trust-related behaviors through 3D motion capture technology and gesture recognition algorithms. We aim to eventually create a hierarchical system—with low-level gesture recognition for high-level trust recognition—that is capable of predicting whether an individual finds another to be a trustworthy or untrustworthy partner through their nonverbal expressions.

## Introduction

Honest signals are primitive social signals communicated between people through unconscious behaviors that contain information about our intentions, goals, and values (Pentland 2008). By observing these unconscious behaviors, we can gain insight into how well a date is going (Pentland 2008) and how successfully a deal will be made (Maddux, Mullen, and Galinsky 2007). Such work suggests that since so much of our communication is beyond words that social robots that can understand nonverbal behavior will be considerably more effective communicators.

Robots have an immense potential to help people in domains such as healthcare, and their success heavily depends on their ability to effectively communicate and interact with us. As robots begin to work closely with us, we should consider some important mediating factors that

can affect the outcome of the human-robot interaction. Interpersonal variables like trust, friendliness, engagement, rapport, and comfort should be designed in such a way that is appropriate in different contexts. In a context like healthcare, where robots are being used to collect personal and sensitive information from patients, trust, in particular, is an essential variable to consider (Wilkes et al. 2010). Studies have shown that increased levels of trust facilitates open communication and leads to more information sharing (e.g., Maddux, Mullen, and Galinsky 2007). Thus, when robots gather medical information, more effective communication and information exchange can likely be achieved by establishing an appropriate sense of trust with the patient.

When designing for such interactions, we need to answer how can a robot: 1) **elicit** the appropriate level of trust to an individual (i.e., *the control signal*) 2) **interpret** how much an individual trusts the robot (i.e., *feedback signal*). The work presented here is an effort in answering this second question of how a robot can understand, or detect, whether an individual finds it to be a trustworthy or untrustworthy partner. This preliminary work describes the design and implementation of a computational model for recognizing interpersonal trust in social interactions. By first investigating unconscious signals like gestural cues, we identify how this nonverbal behavior is predictive of later cooperative, or trusting, outcomes. We then design a probabilistic graphical model around those resulting predictive cues, and through 3D motion capture technology, we attempt to automate this recognition of trust by tracking those trust-related nonverbal cues.

## Study 1: Identifying Gestural Cues

We began our investigation with an exploratory study toward finding gestural cues that are predictive of trusting behaviors. The study consisted of two parts. Participants first began by engaging in a social interaction with another participant in a “get-to-know-you” period of 5 minutes. For the second half of the experiment, each participant individually played the “Give Some Game” (a prisoner’s dilemma type game to measure trust in terms of cooperative or uncooperative behaviors represented by the number of exchanged tokens (Van Lange and Kuhlman 1994)) in separate rooms and then answered some questionnaires. A total of 43 dyadic

interactions, or 86 individuals, participated in Study 1 (34 male and 52 female undergraduate students). This experiment produced interaction data, which consists of the raw videos of the experiment, fully video-coded annotations of the participant’s behaviors (face and body), and questionnaires that included items assessing—with a 7-point Likert scale—how much the participant trusted and liked their partner. Please see our prior work (Desteno et al. 2012) for more details.

## Results

We found that a set of four cues—face-touching, crossing-arms, leaning-back, and hand-touching—are together predictive of distrusting behavior. Through a multilevel linear regression analysis, we found that the more a participant exhibited these disengaging cues, the fewer tokens they gave their partner. Also, the more a partner exhibited these disengaging cues to the participant, fewer the token the participant would offer to that partner (Desteno et al. 2012). We also found an additional set of three cues—leaning-forward, arms-in-lap, and having an open-arms pose—that are together predictive of higher levels of trust [ $b = 0.047$ ,  $p < 0.031$ ] in a linear regression analysis (Lee 2011). These analyses suggest that observations of these seven identified cues can help predict an individual’s assessment of their partner’s trustworthiness.

## Modeling Trust Using HMM

By observing these seven cues throughout the duration of the 5-minute social interaction, we seek to accurately predict how much a participant will trust their partner—with the number of tokens a participant decided to give their partner representing how much he/she trusted the partner to play cooperatively. Specifically, we created two Hidden Markov Models ( $HMM_{low}$  and  $HMM_{high}$ ) to determine whether an individual will exhibit low or high levels of trust towards their novel partner; each gestural cue that occurs throughout a social interaction changes the HMM’s hidden state. At the end of the interaction, the prediction is the class corresponding to the HMM for which the gestural observation data’s log-likelihood is highest. By means of a leave-one-out cross-validation, our best result with 1 state for  $HMM_{low}$ , and 5 states for  $HMM_{high}$ , had a recognition accuracy of 69.44% accuracy with a baseline of 63.89% (a gesture-ignorant model which always predict the most common class). Thus, this trust model is capable of differentiating with above-baseline accuracy whether an individual will have low or high trust judgements of their novel partner by observing the sequence of predictive cues he/she emits.

## Study 2: Capturing 3D Gesture Data

Our preliminary model attempts to predict whether a person will trust their novel partner or not in behaving cooperatively in an economic Give Some Game by observing a set of informative gestural cues unfold in a social interaction. Assuming that we can improve on our prediction accuracy, an automated method of detecting those gestural cues of interest—rather than rigorous hand coding—will enable automated as-

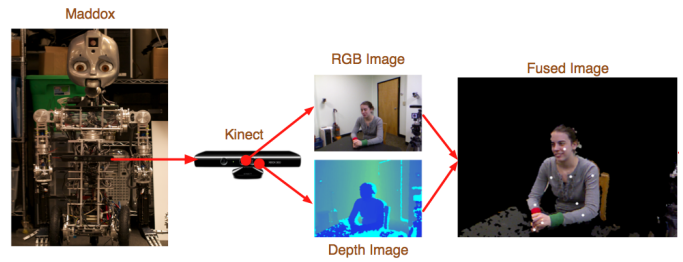


Figure 1: The Kinect provides a depth map and RGB image map of a scene, and OpenNI’s NITE algorithm tracks a person’s movements through skeleton-tracking.

essment of trust. We use 3D motion capture technology and gesture recognition algorithms to detect when these nonverbal cues are being communicated. To gather data, we replicated the scenario in Study 1 but added Kinect sensors in the environment. A total of 28 dyadic interactions, or 56 individuals, participated in this Study 2. By training multiple support vector machines (SVMs), we were able to reliably detect when an individual leans-forward, leans-back, or neither with an average recognition accuracy of 83.7%. However, the remaining five predictive gestures were unreliably detected.

## Ongoing Work

We are currently working to improve accuracy rates for both the trust and gesture recognition systems. We hope to eventually show a hierarchical system—with low-level gesture recognition for high-level trust recognition—that is capable of predicting from nonverbal expressions whether an individual finds another to be a trustworthy or untrustworthy partner.

## References

- Desteno, D.; Breazeal, C.; Frank, R. H.; Pizarro, D.; Baumann, J.; Dickens, L.; and Lee, J. J. 2012. Detecting the Trustworthiness of Novel Partners in Economic Exchange. *Psychological Science*.
- Lee, J. J. 2011. *Modeling the Dynamics of Nonverbal Behavior on Interpersonal Trust for Human-Robot Interactions*. Masters thesis, Massachusetts Institute of Technology.
- Maddux, W.; Mullen, E.; and Galinsky, A. 2007. Chameleons bake bigger pies and take bigger pieces: Strategic behavioral mimicry facilitates negotiation outcomes. *Journal of Experimental Social Psychology* 44(2):461–468.
- Pentland, A. S. 2008. *Honest Signals*. MIT Press.
- Van Lange, P. A. M., and Kuhlman, D. M. 1994. Social value orientations and impressions of partner’s honesty and intelligence: A test of the might versus morality effect. *Journal of Personality and Social Psychology* 67(1):126–141.
- Wilkes, D. M.; Franklin, S.; Erdemir, E.; Gordon, S.; Strain, S.; Miller, K.; and Kawamura, K. 2010. Heterogeneous Artificial Agents for Triage Nurse Assistance. *2010 10th IEEE/RSJ International Conference on Humanoid Robots* 130–137.