Varying Social Cue Constellations Results in Different Attributed Social Signals in a Simulated Surveillance Task

Emilio J. C. Lobato¹, Samantha F. Warta², Travis J. Wiltshire², Stephen M. Fiore²

¹Illinois State University, 100 North University Street, Normal, IL 61761; ²University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816
elobato@ilstu.edu, swarta@ist.ucf.edu, twiltshi@ist.ucf.edu, sfiore@ist.ucf.edu

Abstract

A better understanding of human mental states in social contexts holds the potential to pave the way for implementation of robotic systems capable of more natural and intuitive interaction. In working toward such a goal, this paper reports on a study examining human perception of social signals based on manipulated sets of social cues in a simulated socio-cultural environment. Participants were presented with video vignettes of a simulated marketplace environment in which they took the perspective of an observing robot and were asked to make mental state attributions of a human avatar based on the avatar’s expression of a range of social cues. Results indicated that subtly varying combinations of social cues led to participants’ perception of different social signals. The different mental state attributions made were also significantly associated with what participants considered an appropriate behavioral response for the robot to exhibit in relation to the avatar. We discuss these results in the context of the development of computational-based perceptual systems to be implemented in socially intelligent robots.

Introduction

Robotic systems modeled from human social-cognitive mechanisms hold the potential to not only provide greater insight into human cognitive processes, but may also enable realization of socially intelligent robots (Breazeal, Gray, and Berlin 2009). It is increasingly necessary to understand how humans use information gleaned from interactions to inform their responses to another. Critical to this is the development of social intelligence models that enable robots to translate social cues in the environment, into social signals that allow the robot to interact socially in that environment (Fiore et al., 2013).

In this paper, we report on a study examining the interpretation of mental states to address the question of which social cues people find salient for making specific attributions about the mental states and intentions of others. Research along these lines has recently become a focus for the interdisciplinary field of social signal processing (Vinciarelli, Pantic, and Bourlard 2009) and its application to the context of human-robot interaction (HRI; Wiltshire et al. 2014a). Understanding the social cues that lead to the interpretation of specific social signals, such as threatening or harmless intentions, will provide parameters for the design of more socially capable artificial systems (Fiore et al. 2013). In particular, this will allow for the development of more effective perceptual and cognitive systems for robotic platforms embedded in social contexts. Such contexts are indeed quite pervasive and include any instances in which the robot: (a) observes a person, (b) is observed, or (c) is interacted with. These scenarios are fundamentally social in nature and occur in hospital and caregiver settings, industrial settings, and military contexts.

Social Cognition and Interaction

Social cognition involves the cognitive processes that enable an individual to understand the social environment, including themselves and others (Chatel-Goldman et al. 2013). One major area of social cognitive research is referred to as Theory of Mind (ToM), which concerns an individual’s ability to understand or reason about the mental states of others (Premack and Woodruff 1978; Ibanez and Manes 2012). Specifically, ToM does not just include the process of recognizing mental states, but also the capability to comprehend the factors that comprise a mental state – feelings, thoughts, and intentions – and utilizing them to predict the behavior of others (Uekermann et al. 2006). Mull and Evans (2010) argue that ToM capacity is essential to informing an individual’s ability to comprehend the social environment and engage in meaningful interaction. However, much of ToM-based research does not focus directly on the way in which perceptual cues link to mental states attributed to others (Freeman and Ambady 2011). Our work is guided by the notion that understanding the linkage between social perception and cognition requires a framework that encompasses the social information available in the environment along with an agent’s perceptions and judgments of that information (Wilshire et al. 2014b; Wilshire et al. 2014c).
The framework of social cues and social signals provides such a linkage, which we seek to examine empirically in the present work. Generally, social cues exist in the environment and are expressed by individuals. Social cues are discrete and observable features transmitted via physical or behavioral activity and act as channels of social information that are grounded in emotional, cognitive, social, and cultural contexts. The combination of these cues comprise a social signal conveying the perceived underlying meaning (Fiore et al. 2013; Wiltshire et al. 2014a; Wiltshire et al. 2014b; Wiltshire et al. 2014c).

Both the perception of social cues and the subsequent interpretation of a social signal represent significant aspects of social cognition. Through the mutual exchange of such cues and reciprocal interaction, shared social situation understanding is developed (Pezzulo 2012). The effectiveness of interactive behaviors is largely dependent upon the successful interpretation of another’s intentions (Crick and Dodge 1994; Leffert, Siperstein, and Widaman 2010). For instance, during a social situation, an individual must be capable of correctly interpreting the expressed social cues, determining the intentions of the other agent (e.g., harmless, threatening, friendly), and selecting an appropriate response that facilitates a specified desired result (e.g., deescalating a threatening situation). Social cue and signal relationships are also important for HRI.

Relevance to Human-Robot Interaction (HRI)

In HRI, a robotic platform is conventionally required to convey information to an individual about the surrounding environment. As we, and others, have argued, the environment in which robots are deployed is a fundamentally social environment (Wiltshire, Barber, and Fiore 2013). Therefore, it is crucial that robots are equipped with the ability to perceive and interpret the social aspects of the environment, including the mental states and intentions of others, as well as the appropriate behavioral responses.

The increasing introduction of robots into complex operational environments, such as military operations (Barnes and Jentsch 2010), raises the likelihood that such scenarios may become more common. For example, autonomous or semi-autonomous robots deployed in a surveillance capacity would be tasked with informing teammates of potentially threatening individuals. Human teammates, in turn, could utilize that information to both inform their subsequent behavior as well as give appropriate commands to the robot (e.g., engage or pursue the target to allow for continued surveillance). More generally, such research can inform the development of capabilities in robots to convey social cues indicative of mental states and intentions (e.g., Atkinson and Clark 2013; Klein et al. 2004).

The realization of this scenario, we argue, requires understanding the relationship between social cues and social signals (Fiore et al. 2013; Wiltshire et al. 2014b). One avenue for understanding this complex relationship is to study the social signals that humans report attributing given their observation of certain social cues. The current study aimed to examine this issue via the use of a simulated socio-cultural context where social cues were manipulated. Wiltshire et al. (2014b) proposed a Brunswikian Lens Model approach (Cooksey 1996; Doherty and Kurz 1996; Hammond 1993; Vicente 2003) as an analytic technique to examine the relationship between social cues and social signals. We empirically examine this theorizing as the Lens Model allows for identification of the relationship between social cues and the perceived social signal. By determining weighted levels of activation required for a particular cue or cue combination, the Lens Model approach leads to a probabilistic representation of the cues and their relationship to a specific social signal (Wiltshire et al. 2014b). The long-term goal of this work is provide a foundation for the design of socially intelligent artificial systems capable of interpreting the social environment.

Research Questions

The current study investigated the relationship between social cues and the resulting social signals. The following research questions guided our study: (Q1) Would perception of the social signals attributed to a human avatar in a simulated environment change as a function of differences in the conveyed social cues? (Q2) Would the instructions an individual chose to give a robot vary as a result of differing social signals attributed to the human avatar? Given that this was an exploratory study, our aim was not to try to predict that certain cues would lead to specific signals or that specific signals would result in the administration of certain instructions.

Method

Participants

Participants consisted of 38 students (23 female, 13 male, 2 unreported, \(M_{\text{age}} = 18.44\) years, \(SD_{\text{age}} = 1.3\) years) with the majority identifying as Hispanic/Latino (57.9%). Participants were recruited through a research participation system at a large Southeastern university, and were compensated with course credit.

Materials

The current study took place in a laboratory environment in which participants completed several measures and were presented visual stimuli through use of a computer. All data were collected through the Qualtrics (http://www.qualtrics.com/) system. We created 16 short vignettes using MovieStorm 3D animation and movie-making software (http://www.moviestorm.co.uk/). Each
vignette was 10-20 seconds long and involved human avatars navigating around a Middle Eastern marketplace environment, in which participants were told that a joint task force of United States military and local law enforcement had set up a surveillance robot. The video clips participants watched were shown from the robot’s point-of-view (see Fig. 1). Within each vignette, a pre-selected number of social cues were manipulated for the target avatar. The cues manipulated in the present study were: facial expression, gaze behavior, walking speed, and hand gesture. A novel measure was designed for the current study to allow participants to indicate the: (1) specific affective and cognitive states attributed to the avatars, (2) social cues expressed by the avatars that were most salient in their attributions, and (3) instructions participants gave to the robot.

![Figure 1. Screen capture of the simulated environment from the robot's point-of-view](image)

**Design**

**Independent Variables (IVs)**

The IVs were represented by the cues manipulated within each simulated vignette; (1) **facial expression** (two levels: happy or angry), (2) **gaze behavior** (two levels: gaze oriented towards the robot or oriented in the direction of locomotion), (3) **walking speed** (two levels: fast or slow), and (4) **hand gesture** (two levels: presence of a fist pounding an open hand or absence of any gesture). To better illustrate the manipulation of cues that took place, the cue framework for four of our vignettes is outlined as follows: (a) angry, direction of locomotion, slow, fist smack; (b) happy, direction of locomotion, slow, fist smack; (c) angry, toward robot, fast, no fist smack; (d) happy, toward robot, fast, no fist smack. The remaining vignettes were variations of these social cues and levels.

**Dependent Variables (DV)**

The DVs were: (1) **Specific social signals** attributed to the avatars, which were selected with a yes or no response. The **social signals** listed were: acting suspicious, angry, anxious, distressed, excited, happy, nervous, preoccupied, sad, and threatening. (2) The determination of **social cue salience** for the attributions made was answered with a yes or no response that corresponded to whether the cue was utilized by participants to make the mental state attribution. The **social cues** listed were: posture, head movements, hand movements, facial expression, eyes, movement speed, gait, and clothing. (3) The **instructions** for participants to give the robot were: continue monitoring the scene, pursue the individual while maintaining a safe distance, or pursue and confront the individual.

**Procedure**

Participants were brought into the research laboratory, provided with an informed consent document to review, and allowed to ask questions before electing to continue. Participants were informed that the purpose of the study was to investigate how people use social information to determine the mental states of others. After consent to participate was documented, participants viewed 16 short simulated vignettes of human avatars navigating around a Middle Eastern marketplace environment. After each vignette, participants were asked to make mental state attributions pertaining to the avatar and respond to a scale assessing the social cues deemed salient in making those attributions. Participants were also asked to judge the appropriate instructions that a soldier operating the robot should issue. Lastly, participants completed a demographics questionnaire.

**Results**

Due to space limitations, the present results only detail a subset of the possible attributed social signals. For similar reasons, the data collected were collapsed across all vignettes for analyses, treating the participant responses from each vignette as independent observations. In total, two separate corrected Chi-square analyses were conducted using IBM SPSS Statistics 21. Chi-square tests were utilized given the categorical nature of the data and interest in the relative frequency that certain cues were utilized in attributing a given social signal. Due to the nature of the data collected, the Chi-square assumptions of independence and mutual exclusivity were violated, so a First-Order Rao-Scott Test of Association correction factor was applied (Decady and Thomas 2000). Initial corrected Chi-square analyses examined the salient cues for each social signal participants attributed after viewing the simulated avatar. Secondary corrected Chi-square analyses assessed the association between the social signals attributed and the set of instructions people chose to provide the robot.

**Social Signal x Social Cue**

**Acting Suspicious**

When participants attributed a suspicious mental state to the avatar, they used facial expression (23.2%), hand movement (17.5%), gait (15.2%), movement speed (14.8%), eyes (14.1%), head movement (5.3%), posture (9.1%), and clothing (0.9%). There was a significant association between suspicious mental state attributions
and the types of social cues most salient, $\chi^2_C (8, N = 1968, n_s = 246) = 362.54$, $p < .0001$. People were most likely to use facial expression, then hand movement.

**Angry**
When participants attributed an angry mental state to the avatar, they used facial expression (27.5%), hand movement (20.5%), gait (11.5%), movement speed (9.9%), eyes (16.1%), posture (7.9%), head movement (5.4%), and clothing (1.1%) as social cues. There was a significant association between angry mental state attributions and the types of social cues most salient, $\chi^2_C (8, N = 2696, n_s = 339) = 702.63$, $p < .0001$. People were most likely to use facial expression, followed closely by hand movement.

**Happy**
When participants attributed a happy mental state to the avatar, they used facial expression (40.7%), eyes (15.7%), head movement (14.4%), gait (11.2%), posture (11%), movement speed (7.7%), hand movement (4.7%), and clothing (2.8%) as social cues. There was a significant association between happy mental state attributions and the types of social cues most salient, $\chi^2_C (8, N = 1448, n_s = 181) = 547.83$, $p < .0001$. People were most likely to use facial expression.

**Threatening**
When participants attributed a threatening mental state to the avatar, they used hand movement (25.9%), facial expression (21.9%), gait (13.4%), eyes (13.1%), movement speed (10.5%), posture (8.2%), head movement (6%), and clothing (1%) as social cues. There was a significant association between threatening mental state attributions and the types of social cues most salient, $\chi^2_C (8, N = 1856, n_s = 232) = 365.61$, $p < .0001$. People were most likely to use hand movements, followed by facial expression.

**Lens Models**
Lens Model visualizations were created for the four social signals with the strongest significant effect sizes (see Figures 2-5). Cramer’s V (Cramer 1999) is the effect size used here to represent the strength of association between the mental state and cues, due to both variables exceeding the limit of two categories (Field 2013).

**Robot Instructions x Social Signal**
There was a significant association between the instructions people chose to give to the robot and the social signals attributed. Those who instructed the robot to continue to monitor the scene [$\chi^2_C (10, N = 2617, n_s = 263) = 359.92$, $p < .0001$] were most likely to make a happy (27.4%) or angry (18%) mental state attribution. Those who instructed the robot to pursue and maintain a safe distance [$\chi^2_C (10, N = 2733, n_s = 277) = 978.93$, $p < .0001$] were most likely to make an angry mental state attribution (24.2%), followed by suspicious (20.2%) or threatening (19.3%). Those who instructed the robot to pursue and confront the individual [$\chi^2_C (10, N = 681, n_s = 106) = 411.87$, $p < .0001$] were most likely to make an angry (28.1%) or threatening (24.1%) mental state attribution.
In order to illustrate how our results were calculated, Table 1 provides the positive response frequencies, number of participant observations (n.), uncorrected Chi-square statistic, and correction factor required to reproduce our calculations. These are just an example related to the analyses concerning the social cues mapped to suspicious attributions and the social signals that elicited participants to instruct the robot to continue monitoring the scene.

Table 1. Chi-square Response Frequency Data

<table>
<thead>
<tr>
<th>Social Cues</th>
<th>Acting Suspicious Cue Salience</th>
<th>Robot Instructions to Monitor</th>
<th>Social Signals</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posture</td>
<td>64</td>
<td>Count</td>
<td>Acting Suspicious</td>
<td>33</td>
</tr>
<tr>
<td>Head Movements</td>
<td>37</td>
<td>Count</td>
<td>Angry</td>
<td>80</td>
</tr>
<tr>
<td>Hand Movements</td>
<td>123</td>
<td>Count</td>
<td>Anxious</td>
<td>38</td>
</tr>
<tr>
<td>Facial Expression</td>
<td>163</td>
<td>Count</td>
<td>Distressed</td>
<td>50</td>
</tr>
<tr>
<td>Eyes</td>
<td>99</td>
<td>Count</td>
<td>Excited</td>
<td>51</td>
</tr>
<tr>
<td>Movement Speed</td>
<td>104</td>
<td>Count</td>
<td>Happy</td>
<td>125</td>
</tr>
<tr>
<td>Gait</td>
<td>107</td>
<td>Count</td>
<td>Nervous</td>
<td>10</td>
</tr>
<tr>
<td>Clothing</td>
<td>6</td>
<td>Count</td>
<td>Preoccupied</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Total Responses</td>
<td>Total Participant Observations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>703</td>
<td>246</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Uncorrected Chi-square = 309.62, p < .0001
Correction factor = .854

Uncorrected Chi-square = 297.79, p < .0001
Correction factor = .8274

Discussion

Theoretical Implications

This study examined the relationship between social cues and social signals and provides an initial examination of which social signals were reported for a given set of social cues, and which social cues were considered most salient in making a specific mental state attribution. Q1 sought to determine whether varying the expression of social cues would also vary the perception of social signals attributed to the avatar. We found that the attributions of various social signals were influenced by different social cues and that social cue salience varied for these mental state attributions. For instance, hand movement was most important to attributing an angry or threatening social signal, whereas facial expression was more important in making a happy or suspicious attribution. Our results provide initial evidence regarding the social cue and signal relationship, which has additional implications for future research on social cognition in HRI. Additionally, the methodology we developed furthers understanding of the relationship between combinations of social cues and social signals; however, future research should examine the dynamic and interactive relationship of social cues (Freeman and Ambady 2011).

The present study also investigated how individuals use social signals in determining the appropriate behaviors a robot should be instructed to execute. Specifically, Q2 addressed whether the instructions an individual chose to provide a robot, in regard to the appropriate behavioral response, could vary in accordance with differing mental state attributions. The results suggest that different mental state attributions made by participants led to different behavioral decisions. For example, happy or angry mental state attributions were most likely to result in participants instructing the robot to continue monitoring the scene. Attributions of an angry or threatening mental state were most likely to lead to decisions to confront the individual, and attributions of anger or suspicious or threatening intentions led to more decisions to follow the individual at a safe distance. Whether or not these would be considered the “correct” behavioral response, it is important to keep in mind the behavioral responses that people deem appropriate for a given social situation. While this study was focused largely on behavioral responses for a specific social situation (i.e., a simulated military surveillance operation in a Middle Eastern marketplace), future research could incorporate such behavioral decision trees in other situations to more fully examine the degree to which characteristics of social situations are likely to affect the interpretation of social cues (Wiltshire et al., 2014c).

Practical Implications

In addition to theoretical implications for the study of social cognition, this study yields practical implications for the design of artificial social intelligence systems. For example, our findings can guide the development of bio-inspired artificial intelligence systems (Franklin et al. 2013) seeking to approximate a biological example of how to perform cognitive functions such as social intelligence. In particular, the results of the present study provide a foundation to create models that enable simulation of the attributions for specific social signals given input of particular social cues. Such a system could be programmed to recognize and weight social cues differentially depending on the set of social cues being expressed in order to probabilistically estimate which social signals are being conveyed. With the associated cue utilizations for the constituent social cues, these models could be implemented into an artificial social intelligence system for making mental state attributions.

Progress in this area can be aided by future examination and analysis of additional constellations of social cues in different contexts. Further, this should be conducted with a larger and more representative sample to understand how perceivers differ in attributing social signals from social cues. Likewise, what participants considered an appropriate behavior in response to the attribution of a given mental state could also help advance the development of socially intelligent robots that serve the needs of human team members. For example, if humans are more likely to consider it appropriate to confront individuals displaying a certain combination of social cues over others, robotic systems could be programmed to emphasize or de-emphasize that option when asking for
input from team members after sharing information about the probabilistic mental state of an observed individual.

**Conclusion**

This study examined the role of a small subset of potential social cues in differing mental state attributions during interaction within a simulated socio-cultural environment. Our results also speak to the relative importance, as measured via participants’ self-report, of the perception of an underlying social signal in specifying differentially appropriate robotic behavioral responses. These results have important implications, not only for the design of socially intelligent artificial systems, but also for understanding social cognition more generally. Nonetheless, there is a clear need for further research examining the influence and salience of other social cues, including, for example, non-verbal auditory cues (i.e., intonation), to better understand the complex relationship between cues and signals. Finally, social cues do not exist in isolation, necessitating research efforts that examine, not just social cue combinations leading to the perception of a given social signal, but also the role of larger contextual factors in making mental state attributions.

**Acknowledgments**

This work was partially supported by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0016 (Stephen M. Fiore, PI, sfiore@ist.ucf.edu, corresponding author). Views contained here are of the authors and should not be interpreted as representing official policies, either expressed or implied, of the Army Research Laboratory, the U.S. Government or the University of Central Florida. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation.

**References**


