Preserving Dignity in Patient Caregiver Relationships Using Moral Emotions and Robots

Ronald C. Arkin
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA, U.S.A. 30332

Matthias Scheutz and Linda Tickle-Degnen
Dept. of Computer Science and Occupational Therapy
Tufts University
Medford, MA, U.S.A. 02155

Abstract—This paper provides an overview of an ongoing NSF project that is intended to improve the long-term quality of care for patients suffering from early stage Parkinson’s disease. Due to facial masking in the patient, a stigmatization often occurs in their relationship with the caregiver due to the inability to adequately perceive the patient’s emotional state. We are developing robot co-mediators to assist in the transmission of this information through the use of artificial moral emotions that trigger suitable proxemic expressions in the robot, with the goal of ensuring that the quality of the patient-caregiver relationship is preserved.

Keywords—Parkinson’s Disease, moral emotions, human dignity, intelligent robots

I. INTRODUCTION

As robotics moves toward ubiquity in our society, there has been little concern for the consequences of this proliferation [27]. Robotic systems are close to being pervasive, with applications involving human-robot relationships already in place or soon to occur, involving warfare, childcare, eldercare, and personal and potentially intimate relationships. Without sounding alarmist, it is important to understand the nature and consequences of this new technology on human-robot relationships. To ensure societal expectations are met, this requires an interdisciplinary scientific endeavor to model and incorporate ethical behavior into these intelligent artifacts from the onset, not as a post hoc activity.

Our laboratory has pioneered research in the application of ethics in robotic systems, including the development of an ethical governor capable of restricting lethal behavior [6], and an ethical adaptor that uses the artificial emotion of guilt to affect the available behavioral repertoire consistent with moral restraint [7].

There remains much more to be done however, and not just for the military. While there exists a nascent machine ethics research community, it has focused largely on developmental ethics, often from a purely philosophical perspective, where an agent develops its own sense of right and wrong in situ. In general, these approaches largely ignore the moral emotions as a scientific basis worthy of consideration (e.g., [1]). There is little fundamental or pragmatic research being conducted to ensure that the intelligent robotic systems we create can model and understand our own condition, respond to it in an ethical manner, and preserve the dignity of those people it interacts with.

Currently, there is surprisingly little work on mechanisms that will allow robots to behave in a manner that is ethical and sensitive to the moral context and social norms. This is particularly worrisome as simple robots are already entering society without any notion of ethically acceptable behavior and this situation will only be exacerbated in the future if various kinds of social and assistive robots will cause humans to form unidirectional emotional bonds with robots without those robots being sensitive to human emotions and feelings [24]. In this NSF funded project involving researchers at both Georgia Tech and Tufts University, we tackle a hitherto completely overlooked ethical aspect of human-robot interaction: maintenance of human dignity and the stigmatization of human patients. Unfortunately, in certain health conditions such as early stage Parkinson's disease (PD), it has been shown that the patient caregiver relationship is at risk of deterioration due to disease symptoms that can create inaccurate and negative impressions of the patient’s character. This stigmatization is a serious problem in ensuring the dignity and quality of care that affects both parties in this dyadic relationship.

We are studying ways to minimize the opportunity for such stigmatization to occur. The overarching scientific goal of this project is two-fold: (1) to develop a robotic architecture endowed with moral emotional control mechanisms, abstract moral reasoning, and a theory of mind that allow robots to be sensitive to human affective and ethical demands, and (2) to develop a specific instance of the architecture for a robot mediator between people with “facial masking” due to PD that reduces their ability to signal emotion, pain, personality and intentions to their family caregivers, and health care providers who may misinterpret the lack of emotional expressions as disinterest and an inability to adhere to treatment regimen, resulting in stigmatization [34]. Specific questions we are addressing include: (1) How can an expanded set of moral emotions, particularly empathy, be modeled and exhibited by robots to provide quantitatively better care of patients, in particular, early patients with PD. (2) How can we develop a theory of mind of both caregiver and patient (including their goals and emotional states) that can be used by a robotic mediator to improve the quality of care for patients while enhancing the dignity of both patient and caregiver?

This research is supported by the National Science Foundation under grant #IIS-1317214.
To date, there is little fundamental or pragmatic research being conducted to ensure that the intelligent robotic systems we create can model and understand our own condition, respond to it in an ethical manner, and preserve the dignity of people interacting with those systems. This is, however, especially important for co-robots that are supposed to interact with physically and mentally challenged populations in healthcare settings. In particular, Parkinson’s patients present a large target group where co-robots with moral cognitive mechanisms could make a significant impact both in terms of alleviating pressure on health care providers, improving the interactions between them and their patients, and reducing patient stigma and restoring human dignity.

In this paper, we focus on the role of moral emotions in maintaining patient dignity, leveraging our earlier work in ethical robotic architectures, specifically the use of an ethical adaptor to model and regulate interactions between a robot and a human patient-caregiver dyad.

II. DIGNITY IN HUMAN-ROBOT INTERACTION

Ethical treatment of physically and mentally challenged populations (such as the war wounded, the elderly, children, etc.) is a crucial adjunct to technological advancement in their care. We must not lose sight of the fundamental rights human beings possess as we create a society that is more and more automated. Our research is intended to address fundamental aspects of dignity in human-robot relationships [20]: encroachments on people’s autonomy, maintenance of privacy, freedom from strong or continued pain, and maintenance of self-respect and personal identity. The introduction of robotic technology should not infringe upon these basic human rights. Violation of a person’s dignity results in their degradation, an obviously undesirable effect that could result as these robotic systems are moved into people’s homes, businesses, and institutions. To ensure this does not occur, robot system designers must not only think of the efficiency of their artifacts but also the ethical ramifications of their design choices.

This paper addresses the moral affective aspects of the system. We have extensive experience in the area, having developed a complex time-varying affective model for humanoids as part of our research for Samsung [18] and an architecture for use in Sony’s AIBO and QRIO robots [4,5]. The ethical adaptor component developed in our research for the military has dealt to date with only one of the moral affective functions (guilt) [7]. But there are many others that are far more relevant to the patient-caregiver relationship.

It has been stated that in order for an autonomous agent to be truly ethical, emotions may be required at some level:

“While the Stoic view of ethics sees emotions as irrelevant and dangerous to making ethically correct decisions, the more recent literature on emotional intelligence suggests that emotional input is essential to rational behavior” [1].

Gazzaniga [12] identifies three neuroscientific aspects of moral cognition: (1) moral emotions, which are centered in the brainstem and limbic system; (2) theory of mind, which enables us to judge how others both act and interpret our actions to guide our own social behavior, where mirror neurons, the medial structure of the amygdala, and the superior temporal sulcus are all implicated in this activity; and (3) abstract moral reasoning, which uses many different components of the brain.

These emotions guide our intuitions in determining ethical judgments, although this is not universally agreed upon. Haidt [13] provides a taxonomy of moral emotions: Other-condemning (Contempt, Anger, Disgust); Self-conscious (Shame, Embarrassment, Guilt); Other-Suffering (Compassion); Other-Praising (Gratitude, Elevation). Using the same methods we have in the past [4], we will allow these emotions to bias the behavior of the system in a way which supports a more extensive set of moral emotions, and as appropriate maintain a theory of mind representation of the affective state of the robot’s human counterpart in the relationship in order to act in manner that fosters their emotional state in a manner consistent with enhancing their dignity. Our recent work in this area appears in [37].

III. THE PROBLEM: CAREGIVER-PATIENT STIGMATIZATION

The prevalence of Parkinson’s disease (PD), one of the most common age-related neurodegenerative disorders [39] is characterized by a progressive decline in speed, flexibility, fluidity and coordination of movement throughout the face and body that curtails the person’s ability to non-verbally express feelings, thoughts and intentions to others. People with PD are able to self-report accurately about their interior psychological traits and states, yet their verbal reports are often ignored when delivered with an inexpressive face since the face is one of the most salient and influential mediums of social interaction [34]. The frozen position of pressed or slack lips and an unbroken stare creates the impression, regardless of its accuracy, of an asocial, cold, incompetent or apathetic person who fails to reciprocate others’ feelings of warmth, concern, interest or excitement [9]. The mask interferes with social observers’ formation of accurate impressions regardless of whether the observer is a layperson or a health care practitioner [15,35]. The experience for people with masking is an imprisonment of the self in an unresponsive face and body [3], feeling misunderstood, isolated and lonely [32]. Socially and mentally competent people with masking are confronted with the challenge of clearly presenting a self that overrides observers’ automatic and confidently formed impressions of incompetence [21].

Kurzban and Leary [16] theorize that one of the origins of stigmatization lies in the fundamental social function of dyadic cooperation and the predictability of human behavior for this cooperation to occur. When individuals can quickly decipher one another’s thoughts and actions, they are able to respond effectively as a unit to the task at hand and to predict social and task outcomes. Decipherability is contingent on visible and dynamic physical displays of emotion, thought and motivation. By reducing physical cues of psychological states and traits, facial masking makes it less possible for observers to form reliable impressions of people with PD. When patients
are masked there is a smaller correlation between the patients’ tested social attributes and observers’ impressions of these attributes than when patients are more facially expressive [15]. This indecipherability sets the stage for social response biases and stereotype activation as the social observer attempts to understand an ambiguous situation. The lack of predictability and the reduced perceived trustworthiness of the facially masked patient in comparison to more expressive patients may explain why observers have negative responses to facial masking [35]. People who are more expressive are liked more and viewed as more trustworthy than are less expressive individuals [8]. For example, poor ability to mobilize upper face muscles that produce the eye crinkling of genuine smiling can create a look of deceptive smiling and untrustworthiness in people with PD even though they are experiencing genuine happiness [22]. Consequently, a patient with facial masking enters the health care relationship with, in the words of Goffman [14], a “spoiled identity” created by a special kind of association between attribute and stereotype. The attribute is the muting of facial expressivity and the stereotype is observers’ preconceptions about the meaning of muted expressivity. The health care stigmatization of facial masking has been documented in both western and eastern cultures, yet the quality of the stigmatization is characterized by local moral codes, suggesting that masking violates the implicit moral norms of social interaction [35,40].

When a situation is uncertain, perceivers resolve uncertainty by drawing upon previously formed implicit cultural and normative attitudes and beliefs to form a rapid judgment, after which they become overconfident in the correctness of the judgment and are resistant to revising their initial judgments in subsequent encounters [23,26,31,36]. This overconfidence is augmented by automatic selective attention to evidence that confirms the impression and neglect of disconfirming evidence [23,26]. In general, practitioners have been found to often form incorrect initial impressions of patients’ interior experiences and perspectives, and inaccuracy can compromise the formation of therapeutic alliances and successful interventions. Practitioner accuracy improvements when patients are active rather than passive providers of information in the encounter [31]. Co-robots could facilitate patients’ empowerment as active participants in health care and simultaneously scaffold the practitioners’ utilization of “best practice” methods that are needed to overcome incorrect automatic inferences and biases. Consequently, the co-robot would provide emotional and moral scaffolding for dignified and non-stigmatizing encounters that support the patient’s autonomous contributions to health and the practitioners’ aim of providing intervention that improves the patient’s health.

Specifically, co-robots who “know” the psychological states and traits of their patient, obtained during the comfortable and open exchange of information in the patient’s home, could (1) provide information that alerts the practitioner to consciously develop and analyze alternative explanations to immediate impressions, (2) provide feedback about the practitioner’s accuracy of patient states and traits, and (3) explicitly report data summaries of temporal patterns of health states (e.g., consistency of depressed mood) that help the practitioner to determine the need for intervention (e.g., depression therapy). All of these functions would augment what has been shown in the health research to be needed for practitioners to overcome their incorrect first impressions and biases [26]. These functions are particularly critical for chronic conditions such as PD that have complex and variable symptoms that are difficult for patients to track and self-report in their health care encounters. Furthermore, co-robotic mediation that occurs during the therapeutic encounter may prevent interpersonal self-fulfilling prophecies [30]. By believing highly masked patients to be less psychologically and socially capable than low-masked patients, practitioners may aggravate facial masking symptoms through their bias-confirmatory behavior, thus perpetuating the apparent validity of their initial impressions. They may interview the patient differently based on facial expressiveness and consequently elicit different behavior, such as social withdrawal in the case of higher masking and social engagement in the case of lower masking. For example, Takahashi et al. [33] found that interviewers elicited more positive communication and engagement and less hopeless behavior when they asked adaptive-copying questions of people with PD (e.g., “what did you find satisfying last week? ”) compared to problem-oriented questions (e.g., “what did you find difficult last week? ”). The co-robot might prompt the practitioner to ask adaptive-copying questions during the patient interview in addition to the more typical problem-oriented questions, and by so doing create a more informative, engaging and dignifying encounter.

IV. Approach

Our earlier research yielded a robotic architecture (Fig. 1) capable of supporting a range of ethical activities [6]. We are in the process of expanding the ethical adaptor to incorporate new moral affective functions, such as empathy and compassion and revising and expanding the ethical governor to fit to this new healthcare domain. The relevant ethical components of the architecture are:

1. **Ethical Governor**: A transformer/suppressor of system-generated unethical action to permissible action. This deliberate bottleneck is introduced into the architecture, in essence, to force a second opinion prior to the conduct of a potential morally questionable behavioral response.

2. **Ethical Adaptor**: This architectural component provides an ability to update the autonomous agent’s constraint set and ethically related behavioral parameters and guide action-selection for the robot. It uses a set of moral affective functions at run-time that are invoked in order to ensure proper robotic responses in guiding the patient-caregiver relationship.

While all these components are of value in the patient healthcare context, this research will focus largely on the ethical governor as a means to enforce behavior that supports maintenance of dignity, and the ethical adaptor, which will expand its moral affective inventory to include empathy, in addition to its already present model of guilt [7].
For healthcare applications, we extend this earlier research into a broader class of moral emotions, such as compassion, empathy, sympathy, and remorse, particularly regarding the use of robots for PD, in the hopes of preserving human dignity as these human-robot relationships unfold. There is an important role for artificial emotions in personal robotics as part of meaningful human-robot interaction [5,18], and it is clear that value exists for their use in establishing long-term human-robot relationships. There are, of course, ethical considerations associated with this use of artificial emotions due in part to the deliberate fostering of attachment by human beings to artifacts, and a consequent detachment from reality by the affected user. Most view this as a benign, or perhaps even beneficial effect, not unlike entertainment or video games, but it can clearly have deleterious effects if left unchecked. Hence the need for incorporating models of morality within the robot itself in order to preserve the dignity of the humans involved in these relationships.

![Figure 1: Major Components of the Ethical Autonomous Robot Architecture][1]

V. MORAL EMOTIONS

Finding and creating suitable cognitive models for moral emotions is an ongoing challenge. This secondary class of emotions has been ignored to a high degree by many within the robotics and computer science community, with a few exceptions (e.g., [10]). Once it has been determined that the robot’s actions involve an empathy-inducing situation, it is necessary to compute the appropriate magnitude of empathy that should be expressed. We will use the same methods employed for our models for negative moral emotions [7] to look at positive moral emotions and abstract these into models suitable for computational implementation. The ethical adaptor uses a modified version of the Smits and De Boeck model [29] to compute the system level of a moral emotion.

Using this first approximation, we express a positive moral emotion in terms of situational appraisal values, norm appraisals, other-evaluations, evaluations about the act that elicited the emotion, and motivation and action tendencies geared towards other-support. The model then assigns the probability for “feeling an emotion” as:

\[
\text{logit}(P_{i,j}) = a_j(\sum_{k=1}^{K} \sigma_k \cdot \beta_{j,k} + \tau - \theta_i)
\]

where \(P_{i,j}\) is the probability of person \(i\) feeling the emotion in situation \(j\), \(\text{logit}(x) = \ln(x/(1-x))\), \(\beta_{j,k}\) is the emotion-inducing power of component \(k\) in situation \(j\), \(\theta_i\) is the emotion’s threshold of person \(i\), \(\sigma_k\) is the weight of component \(k\) contributing to the emotion, \(\tau\) is an additive scaling factor, and \(a_j\) is a scaling factor for situation \(j\). In particular, instead of computing the probability that empathy results from some situation, the ethical adaptor computes the magnitude of empathy that robot \(i\) should experience in situation \(j\) as: \(\text{Empathy}_{i,j} = a_j(\beta_{j} - \theta_i)\). In the current implementation of the ethical adaptor, \(\theta_i\) is an initial threshold set for the robot. As above, situational component weights, \(\sigma_k\), ranging from 0 to infinity, represent the relative effect of each component. The additive factor \(\tau\) is derived from user input. Finally, the weight for situation \(j\), \(a_j\), is a scaling factor ranging from 0 to 1 and is related to the necessity of an empathic response.

Creating a computational model is one thing, embedding it into a robotic architecture is another. Fortunately, we have considerable experience in modeling and embedding affect in robotic systems, always based on models derived from cognitive science [4,5,19,25,28]. The infrastructure exists within our MissionLab system [11,17] to be able to facilitate the extension of this approach to lead directly to testing and evaluation.

In order to realize the goals of this research, the ethical adaptor must address three interrelated problems. The foremost of these is the problem of when empathy should be accrued by the system. Empathy, however, does not typically exist in a binary manner, but rather is present in variable amounts. Thus, it is also necessary to determine how much empathy should result from an empathy-inducing action. Finally, it is not enough for the robot to merely accrue empathy from its actions. It is also necessary to define how the ethical adaptor interacts with the underlying behavioral system in order to express its empathy in some manner through behavioral change. In this case, it will be done through the use of non-verbal (kinesics and proxemics) and verbal expressions tailored for the situation.

We use these emotions to bias behavior of robotic systems in human-robot interactions. Our initial focus is on positive moral emotions, and aim to maintain a theory-of-mind of the robot’s human counterpart in the relationship in order to act in manner that fosters their emotional state in a manner consistent with enhancing their dignity. Our recent work in this area, which we intend to expand upon and is related to the use of deception, appears in [38].
VI. CONCLUSIONS

This paper has presented the directions and goals of an ongoing research project into the management of patient-caregiver relationships for early stage Parkinson’s disease. Specific models for moral emotions are posited that can be used for both action-selection and kinesic expression that are hypothesized to lead to improved quality of care.

As this is the beginning of a 5-year project much remains to be done, including the implementation and development of the theory of mind models and their incorporation into the requisite robotic architectures for this work. Human subject studies will ultimately be needed to validate these methods.

REFERENCES