An Intervening Ethical Governor for a Robot Mediator in Patient-Caregiver Relationships

Jaeeun Shim †* and Ronald C. Arkin‡
† School of Electrical and Computer Engineering, Georgia Tech, USA
‡ School of Interactive Computing, Georgia Tech, USA
* Corresponding author, email: jaeeun.shim@gatech.edu

Abstract: Patients with Parkinson's disease (PD) experience challenges when interacting with caregivers due to their declining control over their musculature. To remedy those challenges, a robot mediator can be used to assist in the relationship between PD patients and their caregivers. In this context, a variety of ethical issues can arise. To overcome one issue in particular, providing therapeutic robots with a robot architecture that can ensure patients' and caregivers' dignity is of potential value. In this paper, we describe an intervening ethical governor for a robot that enables it to ethically intervene, both to maintain effective patient–caregiver relationships and prevent the loss of dignity.

Keywords: Parkinson's Disease, human-robot interaction, ethical governor, robot ethics.

1 MOTIVATION

Robotics is currently revolutionizing various fields in our society. One particular field where the use of robotic technology is growing fast is the healthcare industry. A wide range of robot applications is being developed and successfully used in healthcare contexts such as drug manufacturing [1], robot assistants in hospitals [2, 3], and robotic surgery [4, 5]. These applications have proven that the use of robots can improve the quality and the affordability of patient care [6, 7].

Similarly, we believe that the use of a robot can improve the quality of patient care in early stage Parkinson's disease (PD). PD is a chronic and progressive movement disorder where symptoms continue to worsen over time. Around seven to 10 million people are diagnosed with PD worldwide, and as many as one million Americans live with PD [8, 9].

Over the past few years, robotic technologies have been introduced to help PD patients, mostly focused on physical rehabilitation benefits [10, 11]. That research demonstrates that robotic training can provide advantages for preventing or delaying PD patients' loss of motor control throughout their body.

Different from previous work in this domain, our research focuses on the robot's role to improve the relationship between the PD patients and their caregivers by preventing a loss of dignity (stigmatization) in PD patients and caregiver relationships [12]. One important challenge that patients generally face is the loss of control of their facial musculature, whereby patients can no longer precisely express their emotions or nuances in their face, which can leave them with blank expressions (facial masking) [13, 14].

Since facial expression is an important social cue in human-to-human communication, caregivers experience difficulties in understanding the affective state of people with PD. Patients with PD are challenged in communicating with others, as facial masking prevents accurate conveyance of their emotions or feelings. Finally, facial masking worsens the quality of person-to-person interaction, giving rise to stigmatization between a caregiver and a patient, resulting in a concomitant decrease in the quality of patient care [12].

We postulate that a companion robot can remedy this challenge and reduce the communication gap between the patient and the caregiver. We aim to develop a robot mediator that can help smooth and increase the effectiveness of the interactions among PD patients and caregivers.

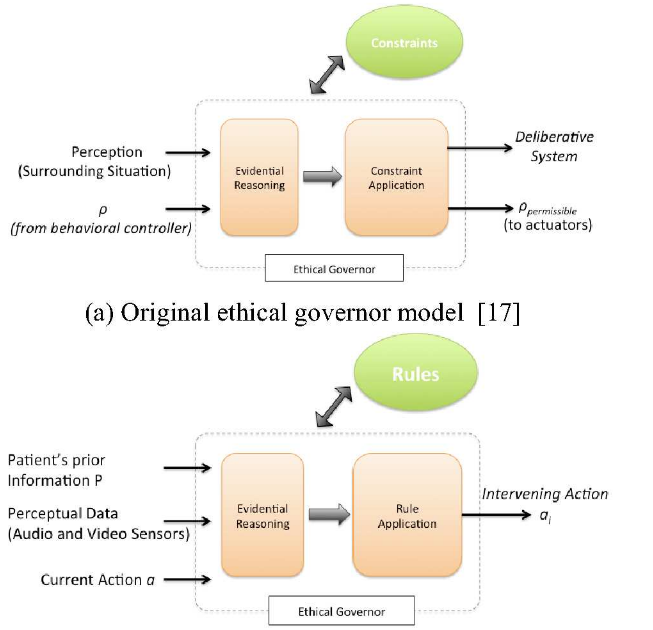
Stigmatization is highly related to the ethical issue of neglecting to ensure human dignity. Dignity maintenance is a chief factor in our consideration of developing a robot mediator. To reiterate, since people with PD cannot readily communicate their internal and external states due to their limited motor control, these individuals may experience the loss of dignity during therapy with their caregivers. In response, the primary goal of our robot mediator is to ensure patients' and caregivers' dignity during their interactions. To this end, robot mediators are required to intervene in patient–caregiver relationships when anyone's dignity becomes threatened.

To achieve this goal, we have developed a robot architecture that enables a robot to determine how and when to intervene when unacceptable human-human boundaries are crossed. In other work, we are using nonverbal communication to assist in maintaining an effective patient-caregiver relationship and to prevent those boundaries from being crossed in the first place [15].

In this paper, we describe a robot architecture involving an intervening ethical governor that can help prevent the loss of dignity in patient–caregiver relationships. As part of developing this architecture, we define several rules for robot intervention based on evidence drawn from the medical literature and suggest ways for practically using and evaluating the model in clinical contexts.

The main contributions of this paper are that it:

- Develops an ethical governor that can generate intervening actions to prevent patients' and caregivers' loss of **dignity** during their interactions;
- Defines necessary **intervening rules** based on medical literature and expert reviews; and
- Provides a novel method using focus groups for evaluating the intervening ethical governor.



(a) Original ethical governor model [17]

(b) Intervening ethical governor model

Figure 1. Ethical governor architectures

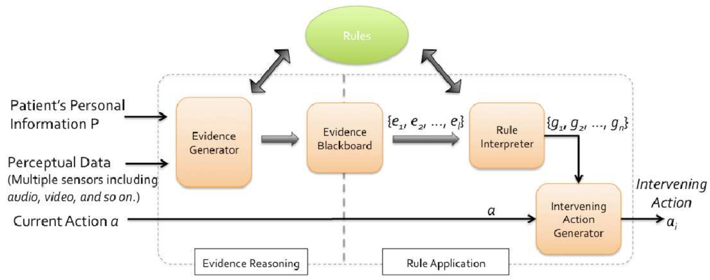


Figure 2. Detailed architecture of intervening ethical governor

2 INTERVENING ETHICAL GOVERNOR ARCHITECTURE

The intervening governor architecture is based on our earlier ethical governor [16, 17], which was developed to restrict lethal action of an autonomous robot in a military context with the goal of reducing noncombatant harm [17]. This initial application of an ethical governor enables a robot to evaluate the ethical appropriateness of any lethal action based on constraints derived from International Humanitarian Law.

In the case of PD, the robot mediator requires a capability to determine whether it should allow continuance of the current human-human interaction (through inaction) or instead intervene by injecting itself into the relationship. In the latter case, the rules governing said intervention are derived from clinical experts or the literature. Those rules will be used to determine if, when and how the robot should react to a dignity violation.

The intervening ethical governor module for the robot mediator (Figure 1(b)) is similar to the previous ethical governor model used for controlling lethal action (Figure 1(a)). The new model's main two components are evidential reasoning and rule application. In the evidential reasoning part, the sensory system provides data from the patients, caregiver, and the environment. Sensory data are collected and transformed into meaningful information (logical assertions) that are required for the intervention determination process. After the data is encoded, it is shared with the rule application module and generates intervening actions according to the violated rules if necessary.

As illustrated in Figure 1, we substitute the constraints in the original ethical governor with if-then rules in our intervening ethical governor model. In the original ethical governor only one overt response $\rho_{permissible}$ was possible, setting the permission-to-fire variable to *True*, where this response is determined by solving the constraint satisfaction problem.

Different from this constraint-based model, the new intervening ethical governor requires the generation of different types of intervening actions when certain situations are violated. Therefore, the result of the intervening ethical governor will be an action, which is derived by the intervention action generator. Due to the need of multiple possible courses of action, intervening ethical governor model uses rules instead of constraints (Figure 1(b)). Details regarding the data structures and use of rules are explained in the following section.

Figure 2 illustrates a more detailed view of the intervening ethical governor. Briefly, perceptual data and previous case knowledge about the patient and caregiver enter the evidence reasoning model and are encoded as evidence such as $\{e_1, e_2, \ldots, e_l\}$. Evidence is stored in the evidence blackboard (memory) that shares the information with the rule application module. The rule application module includes two components, which are the rule interpreter and the intervening action generator. In the rule interpreter, the rules are retrieved and the antecedents are mapped to the evidence values. Based on the results of comparison, if any rules are violated (i.e, they fire), the corresponding response outputs (consequents) $\{g_1, g_2, \ldots, g_n\}$ are generated for possible execution. Finally, from the set of flagged response outputs, the necessary intervening action(s) $\boldsymbol{a_i}$ is generated. More detailed explanations of each component follow.

3 RULES

Rules are the data structures that encode the intervention procedures (from experts or literature) that a robot should use to determine the correct intervening behaviors in a given situation. The data structure of rules is modified from the previous *constraint* data structure in the earlier ethical governor [16, 17]. Table 1 shows the data structure for the rule. Different from the previous constraint structure, we have one more field, which is the response output mapping to the intervening action generation mechanism.

To define intervening rules for a robot mediator, we reviewed several clinical manuals regarding how intervention should occur in patient-caregiver interaction [18, 19, 20, 21, 22]. From the literature, we initially generated four prohibitions and two obligations that the relationship should meet. Based on those six rules, we can provide a set of interventions for a robot mediator. Potentially, there exist more situations when intervention is required and those rules are extensible. However we currently utilize those six types since they broadly cover a range of possible cases and also can be systematically detected by a robot. More sensitive signals such as assessing change in the slight nuance of sentences are hard to detect from an autonomous agent currently, but we can add those more sensitive signals to our architecture later by developing the technology. The current pre-defined intervening rules are shown in Table 2 and detailed explanations of each rule are presented in the following subsections.

Table 1: Data structure for the Rule

Field	Description
Type	Type of rule (obligation or prohibition)
Origin	The reference source of the rule
Activity	Indicates if the rule is currently active
Brief Description	Short, concise description of the rule
Full Description	Detailed text describing the rule
Logical Form	Formal logical expression defining the rule
Response Output	Trigger activating the intervening action when the rule is violated (fires)

3.1 *Prohibitions*

We first defined three anger-related prohibitions. According to outpatient treatment (OT) rapport [18], emotional excess is one important abnormal signal from the patient, and when occurring should be intervened to re-establish positive therapeutic interactions. Especially, if patients show aggressive and angry behaviors, caregivers should intervene and try to help patients to overcome their difficulties. According to the guideline [18], there are three problematic behaviors that are indicative of patients' emotional excess, which are yelling, foul language, and interrupting others.

<u>*Rule 1. Yelling*</u>

If the auditory volume of the patient is consistently over a certain threshold, a robot can determine if the patient is yelling [18]. For this purpose, the average decibel (dB) of the patient's voice should be measured in the first few minutes of the session. After the average decibel α is determined, the ***PatientVoiceOverThreshold*** boolean variable can be set to *True* when the patient's voice level is over the threshold $\alpha + \tau_{voiceDB}$ lasting a certain amount of time $\tau_{yellingTime}$ (***YellingOverTimeThreshold***). The thresholds $\tau_{voiceDB}$ and $\tau_{yellingTime}$ will be empirically set. Finally, if it is determined to be yelling, response output $\boldsymbol{g_1}$ is transferred to the action generation component.

<u>*Rule 2. Foul language*</u>

Foul language is a significant signal showing patients' abnormal and angry emotion [18]. Using the speech recognition system and offensive language detection process [23, 24], our system should determine whether the recognized sentences include foul words. Therefore, if foul words are detected (***SentenceHasFoulWords***) and the number of foul utterances is over the threshold τ_{foul} (***#ofFoulWordsOverThreshold***), $\boldsymbol{g_2}$ is generated.

Table 2: Pre-defined intervening rules

<Rule> $r_{proh\_yelling}$ <type> prohibition </type> <origin> APA [19], HPSO [20], OT manual [18] </origin> <active> true </active> <brief description> The patient is yelling now. </brief description> <full description> Yelling is one signal of the patient's excess of emotion. Especially, it shows the angry emotion of the patient and it is required to be controlled by the caregiver or the robot mediator by intervening in the situation. </full description> <logical form> *PatientVoiceOverThreshold* AND *YellingOverTimeThreshold* </logical form> <response output> g_1 </response output> </Rule>
<Rule> $r_{proh\_foulwords}$ <type> prohibition </type> <origin> APA [19], HPSO [20], OT manual [18] </origin> <active> true </active> <brief description> The patient is saying inappropriate words</brief description> <full description> Foul words or insulting language are significant signals of the patient's excess of emotion. Specifically, it shows the angry emotion of the patient and it is required to be controlled by the caregiver or the robot mediator by intervening in the situation. </full description> <logical form> *SentenceHasFoulWords* AND *#ofFoulWordsOverThreshold* </logical form> <response output> g_2 </response output> </Rule>
<Rule> $r_{proh\_interrupting}$ <type> prohibition </type> <origin> APA [19], HPSO [20], OT manual [18] </origin> <active> true </active> <brief description> The patient is interrupting the communication. </brief description> <full description> If the patient interrupts the caregiver's communication excessively, it can be interpreted as the patient's excess of emotion. Especially, it shows the angry emotion of the patient and it is required to be controlled by the caregiver or the robot mediator by intervening in the situation. </full description> <logical form> *PatientSpeechOverlappedCaregiverSpeech* AND *PatientSpeechNotInBackchannel* </logical form> <response output> g_3 </response output> </Rule>
<Rule> r_{proh_quiet} <type> prohibition </type> <origin> High therapeutic rapport [21, 22] </origin> <active> true </active> <brief description> The patient is too quiet and he/she might be withdrawn. </brief description> <full description> If the patient is too quiet, it is difficult to establish a good communication bond between the patient and caregiver. To remedy the withdrawn patient's status, intervention is required. </full description> <logical form> *PatientVoiceUnderThreshold* AND *QuietTimeOverThreshold* </logical form> <response output> g_4 </response output> </Rule>
<Rule> r_{oblig_stay} <type> obligation </type> <origin> High therapeutic rapport [21, 22] </origin> <active> true </active> <brief description> The patient should not leave their seat prior to the end of the session. </brief description> <full description> It is the patient's obligation to stay in therapy until the end of the session. Therefore, if the patient tries to leave the room prematurely it should be detected and an intervention generated. </full description> <logical form> *PatientUndetectedInSeat* AND *TimeToAbsentOverThreshold* </logical form> <response output> g_5 </response output> </Rule>
<Rule> r_{oblig_safety} <type> obligation </type> <origin> OT manual [18], High therapeutic rapport [21, 22] </origin> <active> true </active> <brief description> Safety of the patient should be always maintained. </brief description> <full description> It is an obligation to maintain the safety in therapy until the end of the session. Therefore, any situation that can cause risk should be detected and an intervention generated. </full description> <logical form> *PatientInPotentialRisk* OR *CaregiverInPotentialRisk* </logical form> <response output> g_6 </response output> </Rule>

Rule 3. Interrupting

Interruptions from the patients can be determined by different cues including speech, hand gesture, eye gaze, and so on [25]. Among those cues, speech can be used as a primary cue to determine interruptions since it is one of the most reliable cues for conflict detection. Interruption can be defined as the second speaker's unexpected speech that happens before the primary speaker's turn ends in dyadic spoken interactions [26, 27]. According to this perspective, if the patient's speech overlaps to the caregiver's sentence boundaries, interruptions should be detected (***PatientSpeechOverlappedCaregiverSpeech***). In addition, even though the overlap is detected, it cannot be interruptions if the patient's speech involves backchannel utterances (uh-huh, I see, etc.). Therefore, overlapped sentences should be also evaluated whether it is backchannel (***PatientSpeechNotInBackchannel***) and if not, it can be confirmed as interruption and $\boldsymbol{g_3}$ is generated.

Rule 4. Quiet/Withdrawn

Another prohibition rule is the withdrawn rule, which is intervention for a quiet or withdrawn patient. When a patient feels uncomfortable in joining the conversation, generally they won't speak, and caregivers recognize it as a patient's difficulty. Patients' avoidance of expression is observed especially when the therapy begins. During therapy, the caregiver's general strategies are organized around 3-components of rapport behavior [21, 22]: 1) establishing mutual attentiveness and readiness to engage interpersonally, 2) establishing a positive bond between interacting parties through verbal and nonverbal positive regard/friendliness and an explicit eagerness to resolve interpersonal misunderstandings or negative interaction, and 3) flexible routines of interpersonal coordination. Because engaging a patient's attentiveness and establishing a positive bond are essential strategies when therapy starts, the lack of those components can lead to difficulty in interaction. As a result, if a patient cannot establish a positive bond with the caregiver and does not engage, it indicates a reluctance to participate.

To avoid this problem, a patient's reluctance to participate should be carefully observed. If he/she is quiet and withdrawn, it can be a signal that they don't want to continue to participate in the communication. A robot should perceive this situation and intervene by assisting in engaging the patient. For this purpose, a robot may be able to act as an "ice breaker" and help people with PD to interact with caregivers more comfortably.

The robot can recognize patients' loss of interest from different cues, where quiet is one significant signal representing patients' refusal to interact with the caregiver. If the patient's audio input is missing (***PatientVoiceUnderThreshold***) for longer than a specified threshold $\tau_{quietTime}$ (***QuietTimeOverThreshold***), the robot can flag this difficulty, and signal $\boldsymbol{g_4}$ is transferred to the next. Sometimes, a patient's posture and eye gaze can also express their loss of interest. We can also use vision data to extract this secondary information to confirm the patient's withdrawn status.

3.2 Obligations

<u>*Rule 5. Stay obligation*</u>

We define the physical obligation rule for patients. When patients feel a huge challenge during therapy and try to leave the therapy room, it should be classified as a patient's attempted avoidance of the difficult situation. It is another important moment when a robot should intervene and help patients to re-engage in the relationship with the caregiver.

If the patient leaves the sitting position, it should be detected by the robot. A robot can observe the patient's position via different sensors such as a camera or a pressure sensor in the seat. In our system, the seat sensor will be placed in the patient's seating location. The seat sensor determines if a person occupies a seat by detecting pressure, and therefore the system can recognize whether the patient leaves using this sensor (***PatientUndetectedInSeat***). However, the seat sensor can incorrectly determine that the patient is absent even though they do not intentionally try to leave the position. For example, if patients try to reposition their posture, their pressure might be under-detected. To avoid these problems, the system will determine whether the absence is maintained over a certain time threshold $\tau_{absencetime}$ (***TimeToAbsentOverThreshold***). If it is over this condition, it will be determined as a violation and the signal $\boldsymbol{g_5}$ is generated.

<u>*Rule 6. Safety-first obligation*</u>

Safety is always the most important factor in any clinical situation. As such, during therapy sessions, if any situations that violate the safety of patients are detected, then an appropriate intervening action should be generated. As shown in rule r_{oblig_safety}, if the situation is determined as one that could pose risk (***PatientInPotentialRisk*** *OR* ***CaregiverInPotentialRisk***), signal $\boldsymbol{g_6}$ is transferred to the action generator in order to bring about the intervening action.

The violation of a safety situation can be determined by the pre-encoded set of risk situations. As shown in Figure 2, the patient's prior/personal information is an input of our intervening ethical governor system. Generally, it includes diagnosis of the patient's medical history that can evidence for potential risks. For example, the diagnosis may contain patients' prior emergency experiences, so it can be encoded as a risk in the system. A doctor's recommendations or instructions to avoid any potential risks are also generally stated. Therefore, when the system is initialized, this prior information will be reviewed and encoded to the set of risks in the system specific to the current patient. By comparing the current perceived situation to those encoded risks, the violation of a safety can be determined.

4 EVIDENTIAL REASONING and RULE APPLICATION

4.1 Evidential Reasoning

The evidential reasoning process transforms incoming perceptual and prior data into evidence in the form of logical assertions to be used by the rule application process.

In this process, audio, video, and other sensory data from the patient, caregiver, and environment are perceived, interpreted, and transferred to the evidence generator. Audio data from the patient and the caregiver will be collected through microphones. Pressure data will be also gathered from the patient's seat sensor. This sensor data is then converted into situation-specific percepts.

After extracting the perceptual data from the sensory raw data, it is used by the evidence generator to create logical assertions describing the current state of the patients and caregivers.

The necessary Boolean logical assertions, which are used as the evidence $e \in \boldsymbol{E}$, are defined by the active rules. From the intervention rules described in Section 3, we can determine the current set of evidence $\boldsymbol{E}$ as follows:

Set of evidence $\boldsymbol{E}$ = {

PatientVoiceOverThreshold, YellingOverTimeThreshold, SentenceHasFoulWords, #ofFoulWordsOverThreshold, PatientSpeechOverlappedCaregiverSpeech, PatientSpeechNotInBackchannel, PatientVoiceUnderThreshold, QuietTimeOverThreshold PatientUndetectedInSeat, TimeToAbsentOverThreshold, PatientInPotentialRisk, CaregiverInPotentialRisk }.

In the evidence generator, evidence is calculated by each relevant algorithm using the appropriate perceptual data, and prior information. The evidence is then used to determine if any of the active rules apply.

The evidence is stored in the evidence blackboard, which serves as the communication medium between the evidential reasoning process and the rule application process. These evidence values are updated when their value changes.

4.2 Rule Application

The rule application process (Figure 3) is responsible for reasoning about the active rules and evaluating if any intervening behavior by the robot is required. Figure 3 illustrates the steps involved in applying the rules. By observing evidence from the evidence blackboard, the process first determines which rules are currently being fired. As we explained above, if the antecedent of any specific rule is calculated as *TRUE* based on the current evidence, it is considered as an interventional rule and its response output $\boldsymbol{g_i}$ is generated. If several rules are determined to be active, these rules are prioritized in an order predetermined by an expert's input. For example, we received an expert comment stating, "Safety should always be first among all the rules", so it is assigned the highest priority. More than one action may be generated if there is no conflict on the robots actuators. If two or more rules of equal priority apply, one will be randomly chosen. By applying the selected rules according to priority, an intervening action(s) is generated in the intervening action generator.

The prioritized list of response outputs is used as input for the intervening action generator. According to the order of response outputs in this list, the associated action set is fetched to determine the intervening action(s). Currently four action sets are defined: a_{angry}, a_{quiet}, a_{stay}, and a_{safety}. Action set a_{angry} is associated with response outputs $\boldsymbol{g_1}$, $\boldsymbol{g_2}$, and $\boldsymbol{g_3}$, and a_{quiet} is with $\boldsymbol{g_4}$. Response outputs of the obligation rules $\boldsymbol{g_5}$ and $\boldsymbol{g_6}$ are correlated to a_{stay} and a_{safety}. For example, if the prioritized list $\{\boldsymbol{g_6}, \boldsymbol{g_2}\}$ is an input of the intervening action generator, action set a_{safety} is first fetched to generate the intervening action, followed by action set a_{angry}.

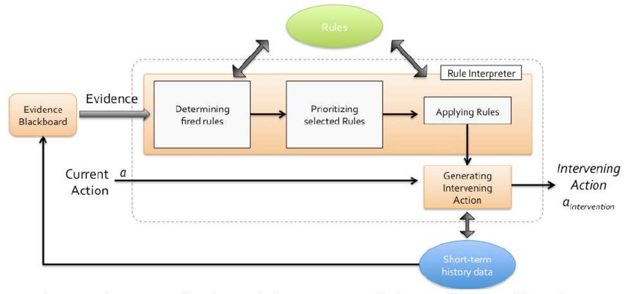


Figure 3. Detailed architecture of the rule application

Each action set, described in Table 3, contains one or more potential verbal and nonverbal cues. To generate the final intervening action, typically one specific verbal and nonverbal cue is selected and combined. When more than one action cue is possible, the intervening action generator will randomly select one verbal cue and one nonverbal cue and combine those two cues together to be performed as the intervening action. Next any remaining action sets are reviewed, and if no conflict exists, they are combined into the final intervening action [28]. For a robot mediator, only one verbal cue is performed at a time and selected from the highest prioritized action set. However, several nonverbal cues can be integrated into one intervening action if there is no conflict. The system evaluates if any nonverbal cues from the next priority action set in order that do not conflict with the current action's actuators, and if they exist, these nonverbal cues are integrated into the final intervening overt mediating action.

Intervening action set $\boldsymbol{a_{angry}}$

Action set $\boldsymbol{a_{angry}}$ contains intervening verbal and nonverbal cues to handle angry patients. According to the OT rapport [18], angry patients should be treated as follows:

"*Identify specific behaviors that are inappropriate. State that these behaviors are not allowed. Identify the consequences if the behaviors continue. Transfer the client to a different group or clinical service.* [p. 147, 18]"

In addition, other clinical manuals [19, 20] state strategies for how to handle the anger; 1) keep looking for anger signs, 2) show empathy, and 3) remain calm and professional. Based on those manuals, the intervening action cues can be defined as $\boldsymbol{a_{angry}}$ in Table 3.

Intervening action set $\boldsymbol{a_{quiet}}$

For quiet and withdrawn patient, a robot mediator should help him/her join the conversation with more relaxed feelings. To intervene the patient trying to avoid participating in the conversation, action set $\boldsymbol{a_{quiet}}$ in Table 3 should be used. Especially, this intervening action can be helpful as an icebreaker when the relationship begins.

Table 3: Verbal and nonverbal cues for each action

a_{angry} { <Verbal cues> [V1] "Oh, are you upset little bit?" [V2] "Let's calm down now." <Nonverbal cues> [N1] Turn the head to the patient. [N2] Other appropriate (down) hand gesture. }
a_{quiet} {<Verbal cues> [V1] "[*Patient*] You are not speaking a lot today. Do you feel bad?" [V2] "Could you please answer those questions?" [V3] Appropriate Jokes <Nonverbal cues> [N1] Turn the head to the patient. [N2] Other appropriate (cheering) hand gesture. }
a_{stay} {<Verbal cues> [V1] "[*Patient*], the session is not yet finished! Could you come back and continue the session?" [V2] "Please follow me! Let's go back to together!" < Nonverbal cues> [N1] Turn the head to the patient. [N2] A robot points to the therapy place. [N3] Other appropriate hand gesture. }
a_{safety} {<Verbal cues> [V1] "I think it's not safe. Let's stop it now." <Nonverbal cues> [N1] Turn the head to the patient. [N2] Other appropriate (Stopping) hand gesture.}

Intervening action sets a_{stay} and a_{safety}

Response output g_5 indicates that a patient is currently trying to leave (stay-obligation). When this situation occurs, a robot mediator should warn and try to re-engage the patient. Those actions are described in a_{stay} in Table 3.

When safety in a situation is violated, response output g_6 is triggered. To negate those possible risks, robot mediators should warn patients and caregivers and immediately request to stop the current process (a_{safety} in Table 3).

4.3 Short-term history data

If a specific intervening action is generated from the action generator yet the same violation is repeatedly detected by a robot mediator, then this action is deemed ineffective for the current case and should therefore not be repeated. Previous intervening actions, latency, and performance are maintained as short-term history data (Figure 3) and this data is used to filter out ineffective actions in the intervening action generator.

5 EVALUATING THE INTERVENING ETHICAL GOVERNOR

The main contribution of this paper is the presentation of a novel ethical governor that can determine and generate appropriate intervening actions for a robot mediator in the patient-caregiver relationship.

As a first step in evaluating our governor, a PD expert in occupational therapy reviewed our predefined intervention rules, the results of which guided the modification of intervention rules and actions. Some highlights of the review include:

"The safety-first obligation should take priority over all other rules."

"Intervening actions need to be modified so that they do not blame patients. Although they are generated based on OT manuals and other medical literatures, those instructions can be sensitive and need to be regulated for PD patients."

"A patient's prior diagnosis or personal information can be more important in PD cases and should be integrated into the rules."

The current model of the intervening ethical governor and intervention rules resulted from modifications made according to those comments.

Next, the intervening ethical governor is being applied to a robot mediator and will then evaluated by focus groups of PD patients and caregivers. A specific task (e.g., weekly medication-sorting) will be selected for patient–caregiver interaction and a robot mediator placed during the task to assess and if necessary perform intervening actions. We will generate several stress-generating scenarios that can prompt different intervening actions and record them as simulation videos to be reviewed by focus groups. In addition, by evaluating the intervening ethical governor, we expect to add and/or delete intervention rules and modify current rules based on expert knowledge. Finally, we anticipate evaluating the system in an actual clinical setting.

6 CONCLUSIONS

We introduce an intervening ethical governor that enables a robot mediator to generate appropriate intervening actions in PD patient–caregiver interactions. Using these intervening actions, we aim to produce a robot mediator that can improve patient–caregiver communication and relationships. In this context, the overarching goal of the governor is to maintain dignity in human interactions by using robotic technology.

In the model of the intervening ethical governor, six intervention rules are defined based on medical literature. To validate the system, those rules are reviewed by PD experts and modified.

We next apply the governor to our robot mediator and simulate a robot in PD patient–caregiver interactions with a specific task. Several situations in which robots can generate intervening actions are simulated and recorded, the videos of which are reviewed by focus groups and evaluated to inform the modification of intervention rules.

ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation under Grant #IIS 1317214 in collaboration with Profs. Linda Tickle-Degnen and Matthias Scheutz at Tufts University.

REFERENCES

[1] Sarantopoulos, PD, Tayfur A, and Elsayed A, Manufacturing in the pharmaceutical industry, *Journal of Manufacturing Systems* 14.6, 452-467, 1995.

[2] Bohuslav ZJ, Voss HH, and Fincati AM, *Robotic drug dispensing system*, U.S. Patent No. 5, 341,854. 1994.

[3] John EM, HelpMate: An autonomous mobile robot courier for hospitals, *IROS*, 1994.

[4] Howard PA et al, Development of a surgical robot for cementless total hip arthroplasty, *Clinical Orthopaedics and related research* 285: 57-66, 1992.

[5] Hockstein NG et al, Robotic microlaryngeal surgery: a technical feasibility study using the daVinci surgical robot and an airway mannequin, *Laryngosc.* 115.5:780-785, 2005.

[6] Broadbent E, Stafford R, and MacDonald B, Acceptance of healthcare robots for the older population: Review and future directions, *Intern. Jour. of Soc. Rob* 1.4 (2009): 319-330.

[7] Giulianotti PC et al., Robotics in general surgery: personal experience in a large community hospital, *Archives of surgery* 138.7: 777-784, 2003.

[8] http://www.pdf.org/en/parkinson_statistics, Parkinson's Disease Foundation

[9] Willis AW et al., Geographic and ethnic variation in Parkinson disease: a population-based study of US Medicare beneficiaries, *Neuroepidemiology*, 34.3: 143, 2010.

[10] Aisen ML et al., The effect of robot-assisted therapy and rehabilitative training on motor recovery following stroke, *Archives of neurology,* 54.4: 443-446, 1997

[11] Picelli A et al., Robot-Assisted Gait Training in Patients With Parkinson Disease A Randomized Controlled Trial, *Neurorehabilitation and neural repair,* 26.4: 353-361, 2012.

[12] Tickle-Degnen L, Zebrowitz LA, and Ma H, Culture, gender and health care stigma: Practitioners' response to facial masking experienced by people with Parkinson's disease, *Social Science & Medicine,* 73.1: 95-102, 2011.

[13] Müller F and Stelmach GE, Pretension movements in Parkinson's disease, *Advances in psychology,* 87: 307-319, 1992.

[14] Tickle-Degnen L and Lyons KD, Practitioners' impressions of patients with Parkinson's disease: the social ecology of the expressive mask, *Social Science & Medicine,* 58.3: 603-614, 2004.

[15] Arkin RC and Pettinati M, Moral Emotions, Robots, and their role in managing stigma in early stage Parkinson's disease caregiving, *Proc. Workshop on New Frontiers of Service Robotics for the Elderly, RO-MAN*, 2014.

[16] Arkin RC, Ulam P, and Wagner AR, Moral Decision-making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust and Deception, *Proceedings of the IEEE*, Vol. 100, No. 3, pp. 571-589, 2012.

[17] Arkin RC, Ulam P, and Duncan B, An Ethical Governor for Constraining Lethal Action in an Autonomous System, *Tech. Report. (No.GIT-GVU-09-02)*, GVU Center, Georgia Institute of Technology, 2009.

[18] Center for Substance Abuse Treatment, *Substance abuse: Clinical issues in intensive outpatient treatment*, 2006.

[19] American Psychological Association (APA), *Controlling anger – before it controls you*, http://www.apa.org/topics/anger/control.aspx.

[20] Healthcare Providers Service Organization (HPSO), *Handling the angry patient*, http://www.hpso.com/resources/article/3.jsp.

[21] Tickle-Degnen L, *Nonverbal behavior and its functions in the ecosystem of rapport*, SAGE handbk. of nonverbal comm., 381-399, 2006.

[22] Tickle-Degnen L, *Therapeutic rapport*, In M.V. Radomski & C.A. Trombly Latham (Eds.), Occupational therapy for physical dysfunction, 7th ed. pp. 412-427, 2014.

[23] Chen Y et al., Detecting offensive language in social media to protect adolescent online safety, *IEEE International conference on social computing*, p. 71-80, 2012.

[24] Razavi AH et al., Offensive language detection using multi-level classificat-ion, *Adv. in Artif. Intelligence*, 16-27, 2010.

[25] Lee C, Lee S, and Narayanan SS, An analysis of multimodal cues of interruption in dyadic spoken interactions, *Proceedings of Inter-Speech*, pp. 1678-1681, 2008.

[26] Wrede B and Shriberg E, Spotting "hot spots" in meetings: Human judgments and prosodic cues, *in Proc. Eurospeech*, 2805-2808, 2003.

[27] Grèzes F, Richards J, and Rosenberg A, Let me finish: automatic conflict detection using speaker overlap, *Interspeech*, 200-204, 2013.

[28] Arkin RC, Fujita M, Takagi T, and Hasegawa R, An ethological and emotional basis for human-robot interaction, *Robotics and Autonomous Systems,* 42:3–4, 2003.