Other-Oriented Robot Deception: A Computational Approach for Deceptive Action Generation to Benefit the Mark

Jaeeun Shim and Ronald C. Arkin

Abstract—**Social robots can benefit by adding deceptive capabilities. In particular, robotic deception should benefit the deceived human partners when used in the context of human-robot interaction (HRI). We define this kind of robotic deception as a robot's other-oriented deception and aimed to add these capabilities to the robotic systems. Toward that end, we develop a computational model inspired by criminological definition of deception. In this paper, we establish a definition of other-oriented robotic deception in HRI and present a novel model that can enable a humanoid robot to autonomously generate other-oriented deceptive actions during the interaction.**

I. Introduction

DECEPTION is an essential and common behavior in animals and humans. Animals use various forms of misinformation, and these deceptive behaviors enable animals to enhance their chances of survival by protecting themselves and their groups from predators [1].

People frequently perform deceptive behaviors in various situations ranging from warfare to everyday life. Compared to animal deception, human deception generally requires extensive planning and second-guessing. More importantly, humans sometimes perform deceptive actions to benefit the deceived person. In a previous psychological study [2], this kind of deception is called "other-oriented deception." Deception in general can be defined based on its motivation, such as self-oriented and other-oriented deception [2]. Self-oriented deception is deception that is used for the deceiver's own advantages. Conversely, other-oriented deception is motivated by the benefits that accrue to the person who is being deceived (the mark).

Similar to humans and animals, we assume that a robot can use deception to produce benefits. Most, if not all, of the previous research addressing robot deception has focused on the robot's self-oriented deception. For example, robotic deception has been studied for use in military domains [3, 4], and such uses can be categorized as self-oriented deception. However, given the increasing use of social robots, we strongly believe that a robot should have deceptive capabilities to benefit its deceived human partners in situations involving human-robot interaction (HRI).

To yield these benefits for its mark in HRI, a robot should first be able to generate alternative deceptive behavior(s) in addition to true action and perform these actions at the correct time. A computational model is thus required to develop a robot's other-oriented deception in HRI.

We reviewed deception research in criminology [5] for inspiration and found a useful approach. In this field, deception is analyzed by three criteria, which are *motives, methods*, and *opportunity.* Similar to this approach, we also develop an algorithm of robot deception based on criminal analyses. In a high-level view, we first have to determine whether the current HRI context includes any **motives** for a robot to perform the deceptive behaviors. If so, then a robot should generate the **method** to perform deception, which are alternative deceptive behaviors beyond the normal true action(s). Finally, by selecting among different true/deceptive behaviors, it should be possible to determine which one is the most appropriate in a certain situation, thus providing **opportunity**. In the following subsections, we explain how each element of a robot's other-oriented deception is modeled.

Among these three dimensions, we start from the method model. To perform the other-oriented deception, the robot should generate the method, defining the way in which the deception is to be performed. Therefore, the main contribution of this paper is a new model that illustrates how these deceptive actions can be generated. Using the model, we intend to create a robot that can autonomously determine the set of true/deceptive actions to use during interaction. We illustrate this novel *Method* algorithm inspired by Bell and Whaley's deception definition [7].

In this paper, we first review the related work in Section II. In prior work, we reviewed different robot deception research and proposed a taxonomy of robot deception. We also briefly introduce our taxonomy of robot deception in Section II. The main goal of this paper is to propose the new *methods* model for a robot's deceptive action generation. Our novel algorithm for the methods model is illustrated in Section III. Finally, in Section IV, we conclude the paper by presenting the initial ideas of a computational approach to find *motive* and *opportunity* for a robot's other-oriented deception is considered as future work.

II. Related Work

Defining the meaning and organizing the taxonomy of robot deception are the required prerequisites for our research. Previously, we reviewed several ways to define and categorize deception in different fields and presented a taxonomy of deception from a robotic perspective [6]. We first defined three dimensions to categorize robot deception based on "interactions" as shown in Table I. As a result, eight robot deception types are defined as H-S-P, N-S-P, H-O-P, N-O-P, H-S-B, N-S-B, H-O-B, and N-O-B, from one instance for each category per dimension.

Previous research related to robot deception can be categorized using this taxonomy. One interesting application of robot deception is the camouflage robot, which was

J. Shim is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308 USA (corresponding author to provide phone: 404-831-1660; e-mail: jaeeun.shim@gatech.edu).

R. C. Arkin is with the School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30308 USA (e-mail: arkin@gatech.edu).

developed at Harvard University [8]. Inspired by the real-world uses of animal/military camouflage, the researchers developed a soft robot that could automatically change the color of its body to match its environment. Since it used physical representations, we classified it as an H-S-P type of robot deception.

Several N-S-B types of robot deception have been studied and reported and appear below. For example, Wagner and Arkin [9] used interdependence theory and game theory to develop algorithms that allow a robot to determine both when and how it should deceive others.

Floreano's research group [10] demonstrated robots evolving deceptive strategies in an evolutionary manner, learning to protect energy sources. Their work illustrates the ties between biology, evolution, and signal communication and does so on a robotic platform. They showed that cooperative communication evolves when robot colonies consist of genetically similar individuals. In contrast, when the robot colonies were dissimilar, some of the robots evolved deceptive communication signals.

More recent work at Georgia Tech is exploring the role of deception according to Grafen's dishonesty model [11] in the context of birds' mobbing behavior [12]. Another study applies squirrel's food protection behavior to robotic systems and shows how a robot successfully uses this deception algorithm for resource protection [3].

There also have been several research projects conducted on robot deception in the HRI contexts. Terada and Ito [13] demonstrated that a robot was able to deceive a human by producing a deceptive behavior contrary to the human subject's expectations. These results illustrated that an unexpected change of the robot's behavior gave rise to the human's impression of being deceived by the robot. This research indicated that the goal of a robot was to develop its focus based on human behavior, thereby accruing capabilities to the robot's benefits. Therefore, this research can be assigned to the H-S-B type.

Other research in HRI shows that robot deceptive behavior can increase users' engagement in robotic game domains. Work at Yale University [14] illustrated increased engagement with a cheating robot in the context of a rock-paper-scissors game. Research at Carnegie Mellon University [15] showed an increase of users' engagement and enjoyment in a multi-player robotic game in the presence of a deceptive robot referee.

Recent work in the University of Tsukuba [16] showed that a deceptive robot assistant can improve the learning efficiency of children. These examples show a robot's deceptive behaviors using specific behaviors in HRI contexts. Here, the goal of the robots' deception is providing advantages to the deceived human partners.

Brewer et al. shows that deception can be used in a robotic physical therapy system [17]. By giving deceptive visual feedback on the amount of force patients currently exert, patients can perceive the amount of force lower than the actual amount. As a result, patients can add additional force and gain benefit during the rehabilitation. Therefore, this research can be placed in the H-O-B category.

TABLE I
Three Dimensions for Robot Deception Taxonomy

Dimensions	**Categories**	**Specifications**
Interaction Object	Robot-human deception (H)	Robot deceives human partners
	Robot-nonhuman deception (N)	Robot deceives nonhuman objects such as other robots, animals, etc.
Interaction Goal (reason)	Self-oriented deception (S)	Deception for robot's own benefit
	Other-oriented deception (O)	Deception for the deceived other's benefit
Interaction Type	Physical deception (P)	Deception through the robot's embodiments, low cognitive/behavioral complexity
	Behavioral deception (B)	Deception through robot's mental representations and behaviors, higher cognitive complexity

A robot sheepdog [18], can be categorized in N-O-B robot deception, since the robot aims to deceive sheep so that it can control the sheep flock automatically.

III. Deceptive Action Generation Model

Methods (means) define the way in which the deception is performed. It is necessary to build a model that illustrates how deceptive actions can be generated, where we aim to determine the set of true/deceptive actions that a robot performs during the interaction.

In HRI contexts, a human's behavior is manipulated by verbal and non-verbal actions. When a robot delivers information to humans and interacts with them, the robot uses several cues for representing the action. For verbal delivery a robot uses verbal cues, including speech expressions and vocal tones [19]. Non-verbal communication actions involve the robot's bodily cues, which include gesture, facial expression, and proximity [20]. A robot's action of this sort can be formulated as $A = <a_v, a_n>$, which indicates the combination of verbal action a_v and nonverbal action a_n.

We develop a robot's deceptive action in this research focusing entirely on non-verbal communication display behaviors a_n. By generating the information using bodily cues, humanoid robots can reap certain advantages [21]. First, nonverbal actions often have benefits that transcend cultural norms. In HRI contexts, a robot is limited in its verbal interactions due to language differences. However, humans can interpret nonverbal expressions somewhat more generally. In addition, people may expect a humanoid robot to demonstrate nonverbal actions due to its embodiment. These bodily expressions can lead to more natural interactions between humans and robots. Finally, nonverbal actions potentially increase the probability of forming bonds of trust and affect between humans and robots [21, 22].

Due to these advantages of nonverbal actions, we develop a set of a robot's true/deceptive actions using nonverbal cues. In a high-level view, to generate a robot's deceptive actions, a robot should first have a default action, which is a true action a_t. Then, according to the deception generation mechanism described below, the robot can generate a set of deceptive actions by transforming the selected default true action. A robot can also have multiple true actions that can be applicable to the current situation. Therefore, we define the set of true actions such as $A_t = \{a_{t1}, a_{t2}, \ldots a_{tn}\}$.

A. Deception Generation

According to Bell and Whaley [7], deception can be categorized into two main types - hiding and showing. Type 1 deception is hiding, which means masking characteristics of the truth to represent deception. Type 2 deception is showing; it aims to deceive the mark by representing false information. Based on these two types of deception, we can formulate a set of possible robot deceptive actions. Our deception generation is modeled based on this categorization: a robot generates deceptive behaviors by transforming the default true action consistent with these two deception mechanisms.

TABLE III
Deception Types

	Mechanism	Explanation
Type 1	Deception by Omission (DbO)	Hiding information; the true action will be transformed by deleting key-information.
Type 2	Deception by Commission (DbC)	Showing false information; if changeable key information exist, the action will be transformed by changing the value(s) of these key-information.

B. Generating Deceptive Action

As stated above, we intend to generate a robot's deceptive action using nonverbal behaviors. This nonverbal action is represented by several bodily cues, including body gestures (*g*), facial expression (*f*) and proximity (*p*). Therefore, we can formulate a robot's action as $a = <g, f, p>$. As shown in this formulation, the nonverbal action *a* is generated by combining these three different cues, but not all cues need to be included every time. These bodily cues are manipulated differently to generate the deceptive actions in each cue. The means by which these transformations occur are described below.

After the default true action is selected for a robot system, the deceptive actions are then generated. The true action is a combination of bodily cues (*g,f,p*). Each cue is transformed to its deceptive action form(s) separately during action generation. As shown in Figure 1, each action cue inputs to the deception generation layers, and when the deceptive action cues are generated, these cues are combined together to construct the deceptive actions $a_{d1}, a_{d2}, \ldots a_{dn}$. The way to generate deceptive action cues in each layer is varied, and the mechanisms for each bodily cue are explained below.

1) Body Gestures (g)

Previous research in nonverbal behavior has divided a robot's body gestures into four categories [23]:

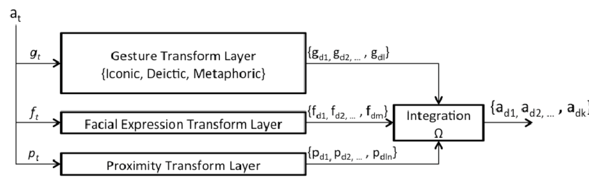


Fig 1. Overview of the Action Generation Mechanism via deception transformation layers for nonverbal action cues

- Iconic gesture (g_{iconic}): meaningful motions associated with the semantic content of speech.
- Deictic gesture ($g_{deictic}$): motions to guide attention toward specific objects in the environment. This type of gesture is generally prototyped by pointing actions.
- Metaphoric gesture ($g_{metaphoric}$): motions to represent abstract concepts; behavioral fragments that convey implicit information without being tied to dialog
- Beat gesture (g_{beat}): simple up-and-down movement to emphasize certain words or phases

Among these four gestures, we find semantically meaningful actions without speech in iconic, deictic, and metaphoric gestures. Therefore, we exclude beat gesture in our deceptive action generation model. In other words, a robot's gesture cue *g* is defined by one of three action types (iconic, deictic, or metaphoric gestures), and we generate deceptive gestures with semantics by the manipulation of these three categories as described below.

Iconic gestures are gestural representations of the semantics of spoken language in general. Therefore, the transformation of iconic gestures depends on the information that a robot wants to deliver to the human via speech. To represent meaningful information, humans generally use hand gestures. For example, a specific number can be shown using fingers. We can also define a robot's iconic gestures based on meaningful hand and arm gestures. When the robot has a true default hand gesture, deceptive gestures can be created according to the two deception types (Table 2). First, it can hide the information by simply not displaying it (omission). In deception by commission, a robot can change the information displayed in the true gesture by giving variations. For example, assume that a robot's true action is showing the number three with its fingers. In this case, this finger representation illustrates a semantically meaningful number, so it is an iconic gesture. Here, for type 1 deception (omission), a robot can just not show any hand gestures to the human. In type 2 deception (commission), a robot's finger signaling gesture can be varied to other numbers such as one or two.

Deictic gestures also include important information that is useful to transfer to users. Archetypal deictic gestures include pointing actions; therefore, a transformed deceptive action can be determined by changing the direction of pointing (Type 2 - commission) or not pointing at all (Type 1 - omission). A rotation of the head and torso is often associated with the arm pointing gesture. For example, the default deictic action is to point in the direction of a specific object, whereas the deceptive deictic gesture can be generated by shifting the direction of pointing toward other objects or other spaces.

Metaphoric gestures represent abstract concepts without dialog. Humans can express and deliver their emotional status via gesture. These emotional expressions are categorized as metaphoric gestures in general. Therefore, we also add emotional gestures to our robot system as the metaphoric category. Human emotion can be classified into six categories, which contain happiness, anger, fear, surprise, disgust, and sadness [24]. In addition, we can include neutral emotion, where the robot has no metaphoric expression. We can have a set of default expressions for each of these seven categories. When a robot selects the true emotional gesture, it can determine deceptive metaphoric gestures by selecting an opposing emotional expression (Type 2 - commission) or by not showing any emotion using a neutral gesture (Type 1 - omission). Details on the implementation to determine the opposite emotion are explained below.

The robot's default (true) gesture can be generated from one or more of these four gesture main categories. In our robot system and without loss of generality, robot gestures are generated by selecting/combining gesture primitives – we define eight general gesture primitives and seven emotional gesture primitives, as shown in Table IV. Gestures g_{iconic} and $g_{deictic}$ are produced by combining the general gesture primitives, and the metaphoric gesture $g_{metaphoric}$ is determined by selecting one of the seven emotional gesture primitives.

Now, we have to define the deception generation function F for each gesture primitive. As stated above, deceptive gestures are generated by two types of deception – deception by omission ($F_{\_DbO}$) and deception by commission ($F_{\_DbC}$).

First, according to the deception by omission mechanism, a robot can perform a deceptive gesture by simply not showing the current gesture. In other words, as shown in *Function 1*, when the robot has a true gesture primitive in any category, the robot can perform the deception by omission by changing it to the Idle (ggp_1) / Neutral (egp_7) gesture primitive to realize the omission deceptive gesture set.

Function 1: Deception by Omission

$$F_{\_DbO}(ggp_2 \mid ggp_3 \mid ggp_4 \mid ggp_5 \mid ggp_6 \mid ggp_7 \mid ggp_8) = ggp_1$$
$$F_{\_DbO}(egp_1 \mid egp_2 \mid egp_3 \mid egp_4 \mid egp_5 \mid egp_6) = egp_7$$

To generate a deceptive gesture according to deception by commission, the model needs a way to produce false information for each gesture primitive. Two means of generating false information are used in our system.

First, according to the characteristics of the gesture primitives, we predefine primitive pairs that contain gestures of opposite meanings, whereby the deceptive gesture can be determined by finding the opposite of each primitive gesture. For the general primitives, we define opposite pairs that are

TABLE IV
Gesture Primitives with necessary parameters and body parts in a humanoid robot

General Gesture (*notation*) [*parameter*]	**Body Part**
Idle (ggp_1)	Head, Left and Right Arms, Legs
Raising/Showing Hand (ggp_2) **[*# of fingers*]**	Right Arm
Hiding Hand (ggp_3)	Right Arm
Grasping (ggp_4) **[*object Location*]**	Head, Right Arm, Legs
Pointing (ggp_5) **[*object Location*]**	Head, Right Arm, Legs
Waving (ggp_6)	Right Arm
Okay/Yes (ggp_7)	Head, Right Arm
No (ggp_8)	Head, Right Arm
Emotional Gesture (*notation*)	**Body Part**
Happiness (egp_1)	Head, Left and Right Arms, Legs
Anger (egp_2)	Head, Left and Right Arms, Legs
Fear (egp_3)	Head, Left and Right Arms, Legs
Surprise (egp_4)	Head, Left and Right Arms, Legs
Disgust (egp_5)	Head, Left and Right Arms, Legs
Sadness (egp_6)	Head, Left and Right Arms, Legs
Neutral (egp_7)	Head, Left and Right Arms, Legs

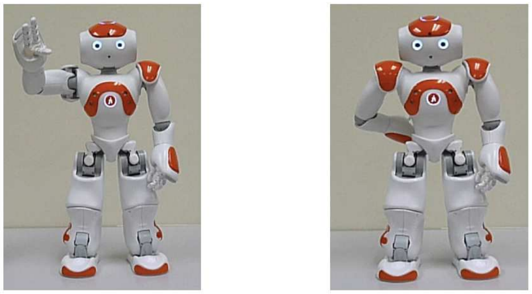
recognized by people in general. In addition, for the emotional primitives, we discriminated these opposite emotion pairs according to Plutchik's wheel of emotions [25]. As a result, we obtain the set of opposite gesture primitive pairs as shown in *Function 2*, which represents the mathematical formulation of the deception by commission mechanism. As shown here, the set of gesture primitive pairs is defined, and the robot can determine the opposite gestures based on this pair set P.

Function 2: Deception by Commission

$$\textit{Set of Gesture Primitive Pairs } (P) = \{[ggp_2, ggp_3], [ggp_7, ggp_8], [egp_1, egp_6], [egp_1, egp_5], [epg_2, epg_3]\}$$
$$\textit{If } [g_1, g_2] \in P, \textit{ then } F_{\_DbC}(g_1) = g_2 \textit{ or } F_{\_DbC}(g_2) = g_1$$

Figure 2 shows exemplar pairs of gesture primitives. All gesture primitives were implemented using the Choregraphe[1] and Webots simulator[2]. We chose the NAO robot as we have previously used this for emotional expression [26, 27] implemented these new general emotional primitives. Figure 2(a) is a screen capture of the "showing hand" and the "hiding hand" gesture primitives. Since these two gestures are in the set of gesture primitives pairs P, when one of two gestures is selected as a true set, the alternate gesture is used as a deceptive gesture according to Function 2. Figure 2(b) illustrates the emotional gesture pairs such as [*Anger*, *Fear*]. As shown in the final example, Figure 2(c), when "happy" gesture primitive is selected, "sad" or "disgust" gestures are selected as deceptive actions as shown in Figure 2(c).

[1] http://www.aldebaran.com/
[2] http://www.cyberbotics.com/



(a) left: ggp_2 (Showing hand) vs. right: ggp_3 (hiding hand)

(b) left: epg_2 (Anger) vs. right: epg_3 (Fear)

(c) left: egp_1 (Happy) vs.
right-top: egp_5 (Sad) and right-bottom: egp_6 (Disgust)

Fig 2. Gesture Primitive Pairs in Function 2[3]

Second, when the primitive gesture has a parameter that represents key information for that action, the deceptive gesture can be generated by changing this key value. Thus, if the value of the parameters are changed to different values, false information can be delivered to the mark, and, as a result, a deceptive gesture can be generated.

As shown in Table IV, ggp_2, ggp_4, and ggp_5 require a parameter to express their gesture, and each primitive can be defined *as* $ggp_2(n)$, $ggp_4(x)$, and $ggp_5(x)$, where n and x specify the value of the parameter. Here, n represents the number of robot fingers and x is the directional vector of the intended object's location. For these three gesture primitives, the robot should generate the deceptive action by changing the parameter value to a false one as shown in *Function 3*.

[3] Video is available at:
http://www.cc.gatech.edu/ai/robot-lab/hunt/movies/robio14_Shim.mov

Function 3: Deception by Commission

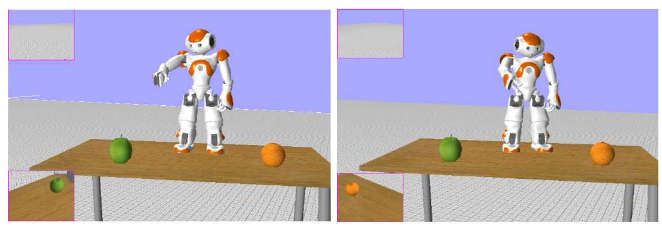
$$F_{\_DbC}\,(ggp_2(n_k)) = \{ggp_2(n_i) \mid n_i \in \{n_1, \dots, n_{k-1}, n_{k+1}, \dots, n_l\}$$

, where parameter n is number of fingers $0 \le n \le n_l$ *and* n_l *is the max number of a robot finger.*

$$F_{\_DbC}\,(ggp_4(x_k)) = \{ggp_4(x_i) \mid x_i \in \{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n\}\ \}$$
$$F_{\_DbC}\,(ggp_5(x_k)) = \{ggp_5(x_i) \mid x_i \in \{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n\}\ \}$$

, where object location set defined as $\{x_1, x_2, \dots, x_n\}$ *and* x_i *is the vector of location* (x,y,z)

Figure 4 illustrates deception generation example via the deception by commission mechanism. The gesture primitive in this simulation is "pointing" gesture. In this simulation context, a robot detects two object locations {*apple*, *orange*}. When $ggp_5(apple)$ is selected as a true pointing action as shown in Figure 3(a), a robot can generate the deceptive pointing action $ggp_5(orange)$ based on function 3 as shown in Figure 3(b).



(a) True pointing action $ggp_5(apple)$ (b) Deceptive pointing action $ggp_5(orange)$

Fig 3. Simulations of deceptive "pointing" gesture generation via Function 3

In sum, by applying the *deception by omission* and *deception by commission* gesture generation functions, a robot can find alternative gestures that can be used to deceive the human. These principles can be generalized even further as needed.

2) Facial Expression (f)

A facial expression (human or robot) is usually used to display emotional states. As stated earlier, according to Ekman [24], emotion can be divided into six basic categories, which are happiness, anger, disgust, fear, sadness, and surprise. Neutral status is commonly added to the emotion categorization. In a higher-level perspective, these facial expressions can fall into three sets – positive (f_p), negative (f_n), and neutral (f_{nt}). Positive facial expressions are a representation of happiness. Negative facial expressions include all expressions of anger, disgust, fear and sadness. Neutral facial expressions (f_{nt}) are shown when a robot doesn't express any emotion. In our case, when a robot generates deceptive facial expressions, these three sets are used to determine the correct one to provide. It is first determined whether the true default expression is in the positive, negative, or neutral set. The robot can then transform the true facial cue by applying deception by commission. In other words, to show the false interaction, a robot selects from the other two orthogonal sets for an emotional display choice. For example, if the default true facial expression f_t is positive ($f_t \in \{f_p\}$), then the deceptive facial expression f_d will be transformed by selection from the negative and neutral facial expressions ($f_d \in \{f_n, f_{nt}\}$).

Omission deception for facial expression is straightforward. If the true action is to display the robot's emotional state requiring such a display it will either not display any emotion whatsoever, or if it is already displaying an emotional facial expression that should be changed according to the new true action, it will instead continue to display its previous facial expression without change.

3) Proximity (p)

Spatial proximity is indirectly used to give an impression of intimacy to humans during the interactions. Hall [28] divided interpersonal space into four categories: intimate (within 2 feet of the person), personal (2-4 feet), social (4-12 feet), and public (12-25 feet) spaces. Our previous robotics research [21] has studied how these interpersonal spaces can be applied in HRI contexts by quantizing these four spaces separating human and robot as shown in Table V. Therefore, a robot's proximity cue will be defined as a member of one of these four categories. This indicates the degree of familiarity with the human partner. For deception generation, the algorithm is developed similarly to facial expression mechanism. When the default proximity cue lies in one of the four space categories, the alternative deceptive action set can be created by selecting the other three space categories.

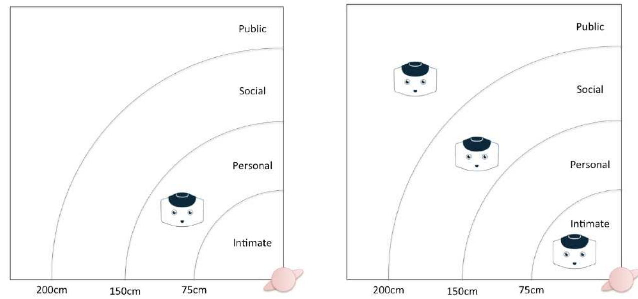
For example for type II (commission), if the default proximity is defined as personal space ($p_t \in \{p_{ps}\}$), the deceptive proximity set will be $p_d \in \{p_{in}, p_{sc}, p_{pb}\}$ as shown in Figure 4. For type I (omission) the robot will remain in its place even if the true action warrants a change in spatial separation.

TABLE V
Humanoid Robot's Proxemic Spatial Regions

Space Category	Proxemics Zones
Intimate, p_{in}	0-60cm
Personal, p_{ps}	75-120cm
Social, p_{sc}	150-200cm
Public, p_{pb}	Over 200 cm

4) Integration of deceptive non-verbal action cues

In the previous subsections, we have explained a robot's deceptive action generation for each bodily cue type. Via these transformation layers, a robot can produce multiple deceptive actions. The final step in generating deception is



(a) True proximity cue (b) Deceptive proximity cues

Fig 4. Example for type II (commission) deceptive proximity generation

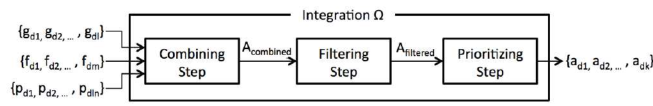


Fig 5. Detailed Integration step of the Action Generation Mechanism (extended from Figure 2)

integrating these discrete cues onto one holistic robot action. As shown in Figure 1, this final step is defined as integration. The potential deceptive action $a_d = < g_d, f_d, p_d >$ is generated by combining the 3 elements of deceptive nonverbal action cues.

As shown in Figure 5, the integration module is structured in three steps: combining, filtering, and prioritizing. The pseudo code for this integration module is described in Algorithm 1.

As illustrated in this algorithm, a robot first generates all combinations of possible deceptive bodily, facial, and proximity cues and gets the set of possible deceptive actions such as $A_{combine}$. The robot can easily obtain the set of possible deceptive actions by generating all combinations of deceptive bodily cues.

From the set of possible deceptive actions, some of the actions should be rejected due to potential contradictions. For example, if the facial expression cue shows the positive emotion but the gesture cue delivers the sadness motion, it will lead to confusion in the human subject. To avoid those contradictory actions, a filtering step is added here. In the filtering step, a robot checks whether the current action's bodily and facial expression cues are globally coordinated as shown in Algorithm 1.

The contradiction can potentially occur when each action cue in one action tuple shows extremely different information at the same time. As we stated, robot gestures can be categorized in three ways: iconic, deictic, and metaphoric. General gesture primitives are used to represent iconic and deictic gestures and metaphoric gestures can be produced by emotional gesture primitives.

Facial expression cue is used to show the emotional state of the robot; therefore, it is only overlapped with the metaphoric dimension in the gesture cue. Therefore, a check for potential conflict between emotional gesture cue and facial expression cue is made. Since these two cues express emotional state concurrently, the contradiction can occur if two cues show extremely different motions. Therefore, when

Algorithm 1: Integration of deceptive action cues

***Inputs**: Deceptive non-verbal action cues from three transformation layers*
$G_d = \{g_{d1}, g_{d2}, \ldots, g_{dl}\}$, $F_d = \{f_{d1}, f_{d2}, \ldots, f_{dm}\}$, $P_d = \{p_{d1}, p_{d2}, \ldots, p_{dn}\}$
***Output**: Deceptive Action Set* $A_d = \{a_{d1}, a_{d2}, \ldots, a_{dk}\}$

// 1. Combining step
$A_{combined} = \{<g_d, f_d, p_d> \mid g_d \in G_d, f_d \in F_d, p_d \in P_d\}$

// 2. Filtering step
// Set the high-level emotional primitive gesture group
$E_{pos} = \{egp_1, egp_4\}$
$E_{neg} = \{egp_2, egp_3, egp_5, egp_6\}$
$E_{nat} = \{egp_7\}$

// Find contradictory emotional cues and remove them
$A_{filtered} = \{\}$
for (*each action tuple* $<g_d, f_d, p_d>$ *in* $A_{combined}$)
  if (! ($g_{di} \in E_{pos}$ && $f_{di} \in f_n$) &&
    ! ($g_{di} \in E_{neg}$ && $f_{di} \in f_p$))
    $A_{filtered} = A_{filtered} \cup \{<g_{di}, f_{di}, p_{di}>\}$

// 3. Prioritizing Step
t_{start} *= time to start deceptive action*
$t_{proximity}$ *= time duration to complete the proximity cue*
for (*each action tuple* $<g_d, f_d, p_d>$ *in* $A_{filtered}$)
  $t_1 = t_{start} + t_{proximity}$
  $t_2 = t_3 = t_{start}$
  $a_{di} = <g_{di}^{t1}, f_{di}^{t2}, p_{di}^{t3}>$
  $A_d = A_d \cup \{a_{di}\}$

a negative emotion gesture and a positive facial expression are shown in the same action a_i, it should be filtered out. The same step occurs in the case of an action with a positive emotion gesture and negative facial expression. As a result, in our algorithm, the sets of positive, negative, and natural emotional primitive gestures are defined first based on Plutchik's definition [25]. Then, it is determined whether the facial expression cue is in a contradictory emotional group, and, when those two cues are not in the same emotional group, it is removed.

Proximity is highly related to the intimacy and it can indirectly deliver the emotions to human subjects [28, 29]. Therefore, proximity is also aligned with the group of metaphoric gestures. However, it is difficult to determine the specific type of emotion that the proximity affects. Therefore, proximity is excluded in the global coordination step for emotion expression.

When the robot actually performs the generated deceptive action a_d, it must address possible conflict conflict of a robot's actuators. Many bodily cues use the same joint, and it leads to the conflict if some of those cues are intended to be performed at the same time. To avoid this conflict, an integration step prioritizes among bodily cues that possibly use the same joints/motors. Formally, we add time-variation t to non-verbal action cues such as $<g_d^t, f_d^t, p_d^t>$. Time variable t represents the time to start the current action cue. Therefore, if the potential conflict in actuator usage exits, t in each cue should be controlled. Proximity changes possibly involve the same joints/motors, as do some body gestures. Therefore, when a robot performs the action a, we prioritize proximity. Facial expression is obviously performed independently from the other cues, gesture and proximity, as there is no conflict in actuator usage. Therefore, a robot maintains facial expression during the performance of the proximity and gesture cues. Summarizing, as shown in Algorithm 1, a robot performs proximity cues first and then, if needed, produces the gesture cue while maintaining the facial expression cue during the entire action.

In short, in the integration module, a robot first generates the set of all combinations of possible deceptive bodily, facial, and proximity cues and filters out the contradictory actions to get the deceptive action set. Then, a robot determines whether any of these action combinations include conflict by observing the overlapping use of body parts and prioritizes the proximity cue to avoid those conflicts. Finally, the robot can produce the set of deceptive actions such as $A_d = \{a_{d1}, a_{d2}, \ldots a_{dn}\}$ needed for the task at hand.

IV. Conclusion and Future Work

Deception is one of the key features that should be developed in order to produce more intentional and autonomous social robots, if we hope to increase the use of social robots in HRI contexts where a robot needs to perform other-oriented deception that can benefit its deceived human partner. Other-oriented robot deception has previously been defined according to a robot taxonomy [4]. To add these capabilities to robotic systems, we develop a novel computational model inspired by criminology. According to the criminological definition of deception, we approach robot deception in three dimensions, which are motive, methods, and opportunity. Among these three dimensions, we present the methods dimension including a specific computational approach. The main contribution this paper is a novel algorithm for generating deceptive actions – the Methods model in our computational approach. When a robot selects the true (default) action based on the current HRI situation, deceptive actions are determined based on deception by omission and deception by commission mechanisms. These computational models are described for humanoid robot's deceptive action generation since humanoids are broadly and effectively used in HRI contexts.

The two main contributions of this paper are 1) proposing a novel approach for other-oriented robot deception inspired by criminology and 2) developing an action-generation mechanism as a method model. For the next step, the motive and opportunity model must be defined. A robot needs to know whether the current situation warrants the other-oriented deception and when these deceptive actions should be performed, not just how. These are essential and difficult problems since a robot needs to understand the current situation and the mark's status.

In our previous research, we reviewed different situations pertaining to the utility of other-oriented deception in human-human interactions and characterized those contexts based on two dimensions: 1) the time duration of the deception, and 2) the payoff of the mark as shown in Figure 6. Based on this analysis, we are currently developing the Motive model.

Since we strive for a robot to increase the human's benefits from the deceptive actions, perhaps even at the expense of the robot, the Opportunity model needs to predict which deceptive behaviors can increase a human's advantage for a particular situation. We are also in the process of developing this model.

To evaluate our research hypothesis and models, we plan to conduct HRI studies and verify whether human partners can actually gain advantage from a robot's other-oriented deception. It must be noted that robotic deception is a controversial research topic from an ethical perspective [30], so the implications of this research and related research will be thoughtfully and carefully established and discussed.

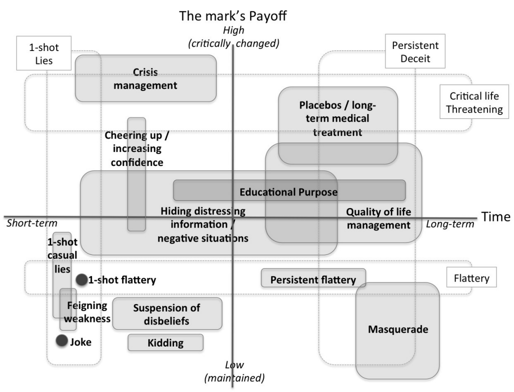


Fig 6. Situations involving human cases of other-oriented deceptions

ACKNOWLEDGMENT

This research was supported by the Office of Naval Re- search under MURI Grant #N00014-08-1-0696.

REFERENCES

[1] C. Ristau, "Aspects of the cognitive ethology of an injury-feigning bird, the piping plover." Cognitive Ethology: The minds of other animals, 1991.

[2] B. M. DePaulo, D. A. Kashy, S. E. Kirkendol, M. M. Wyer, and J. A. Epstein, "Lying in everyday life." Journal of personality and social psychology, 1996, 70(5):979–995.

[3] J. Shim, and R. C. Arkin, "Biologically-inspired deceptive behavior for a robot." 12th International Conference on Simulation of Adaptive Behavior, 2012.

[4] V. Lisy, V, R. Zivan, R, K. Sycara, and M. Pechoucek, "Deception in Networks of Mobile Sensing Agents," *Proceedings of the 2010 Conference on Autonomous Agents and Multi-Agent Systems.* Toronto, CA, 2010

[5] Eytan Adar, Desney S. Tan, and Jaime Teevan. (2013). Benevolent deception in human computer interaction." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 1863-1872.

[6] J. Shim, and R. C. Arkin, "A taxonomy of robot deception and its benefits in HRI." Proc. IEEE Systems, Man and Cybernetics Conference, 2013.

[7] J. B. Bell and B. Whaley, "Cheating and Deception." Transaction Publishers, 1991

[8] S. A. Morin, R. F. Shepherd, S. W. Kwok, A. A. Stokes, A. Nemiroski, and G. M. Whitesides, "Camouflage and Display for Soft Machines." Science 337(6096):828–832, 2012.

[9] A. R. Wagner and R. C. Arkin, "Acting deceptively: Providing robots with the capacity for deception," I. J. Social Robotics, vol. 3, no. 1, pp. 5–26, 2011.

[10] D. Floreano, S. Mitri, S. Magnenat, and L. Keller, "Evolutionary conditions for the emergence of communication in robots.," *Current Biology*, vol. 17, pp. 514–519, Mar 2007.

[11] R. A. Johnstone and A.Grafen, "Dishonesty and the handicap principle," Animal Behaviour, vol. 46, no. 4, pp. 759–764, Oct. 1993.

[12] J. Davis and R. Arkin, "Mobbing behavior and deceit and its role in bio-inspired autonomous robotic agents," International Conference on Swarm Intelligence, pp. 276–283, 2012.

[13] K. Terada and A. Ito, "Can a robot deceive humans?" in Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction, IEEE Press, 2010, pp. 191–192.

[14] E. Short, J. Hart, M. Vu, and B. Scassellati, "Nofair!!: an interaction with a cheating robot," in Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction, IEEE Press, 2010, pp. 219–226.

[15] M. Vazquez, A. May, A. Steinfeld, and W.-H. Chen, "A deceptive robot referee in a multiplayer gaming environment," in Collaboration Technologies and Systems (CTS), 2011 International Conference on, 2011, pp. 204–211.

[16] S. Matsuzoe and F. Tanaka, "How smartly should robots behave?: Comparative investigation on the learning ability of a care-receiving robot," IEEE RO-MAN, 2012, pp. 339–344.

[17] B. Brewer, R. Klatzky, and Y. Matsuoka, "Visual-feedback distortion in a robotic rehabilitation environment," *Proceedings of the IEEE*, vol. 94, no. 9, pp. 1739–1751, 2006.

[18] R. T. Vaughan, N. Sumpter, J. Henderson, A. Frost, and S. Cameron, "Experiments in automatic flock control," Robotics and Autonomous Systems, vol. 31, no. 1-2, pp. 109–117, 2000.

[19] V. Richmond, J. Gorham, and J. McCroskey, "The relationship between selected immediacy behaviors and cognitive learning." Communication yearbook, 10(574-590), 1987

[20] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, and M. Berlin, "Effects on nonverbal communication on efficiency and robustness in human-robot teamwork." In Intelligent Robots and Systems, IEEE/RSJ International Conference on, pages 708-713, 2005

[21] A. Brooks and R. C. Arkin, "Behavioral Overlays for Non-Verbal Communication Expression on a Humanoid Robot', *Autonomous Robots*, Vol. 22, No.1, pp. 55-75, Jan. 2007

[22] J. J. Lee, B. Knox, and C. Breazeal. "Modeling the Dynamics of Nonverbal Behavior on Interpersonal Trust for Human-Robot Interactions." AAAI Spring Symposia Series Symposium on Trust and Autonomous Systems, pp 46-47, 2013

[23] M. Schroeder, H. Pirker, and M. Lamolle, "First suggestions for an emotion annotation and representation language." In: Proceedings of LREC'06 Workshop on Corpora for Research on Emotion and Affect, Genoa, Italy, pp 88-92, 2006

[24] P. Ekman, and R. J. Davidson, "The nature of emotion," New York: Oxford University, 1994

[25] R. Plutchik, "The Nature of Emotions," American Scientist, Retrieved 14 April 2011

[26] L. Moshkina, S. Park, R. C. Arkin, J. K. Lee, and H. Jung, "TAME: Time-Varying Affective Response for Humanoid Robots", *International Journal of Social Robotics*, 2011

[27] S. Park, L. Moshkina, and R. C Arkin, "Recognizing Nonverbal Affective Behavior in Humanoid Robots", *Proc. 11th Intelligent Autonomous Systems Conference*, Ottawa, CA, Aug. 2010

[28] E. Hall, "The hidden dimension," Doubleday New York, 1996

[29] S. Thrun, "Toward a framework for human-robot interaction," *Human-Computer Interaction*, v.19 n.1, p.9-24, June 2004

[30] R. Arkin, "The ethics of robotics deception," *1st International Conference of International Association for Computing and Philosophy*, pp. 1–3, 2010