First the action has to be perceived for communication to take place*

Kiran M. Sabu[1,*], Jennifer Renoux[1], Hermine J. Grosinger[1] and Alessandro Saffiotti[1]

[1]*Center for Applied Autonomous Sensor Systems (AASS), Örebro University, Örebro, Sweden*

Abstract

In this short paper, we consider scenarios in human-robot collaboration where the robot relies on deliberation to activate communicative actions. We claim that reasoning about the perception of these actions is a key, but often disregarded ingredient for successful communication, and propose a pre-theoretical model that accounts for this. We also discuss the problems that may arise if perception is neglected when reasoning about communication.

Keywords

Deliberative Communication, Human-Robot Interaction, Human-Agent Communication, Human-Agent Collaboration.

1. Introduction

In many situations, humans and artificial agents work together to perform joint tasks. Each of them may have their own strengths and weaknesses related to their capabilities. Communication can help the human and the agent in communicating their needs and help each other. For example, in a Human-Robot collaborative setting, the robots may have to act as communicators, and autonomously communicate with their human teammate when the need arises. Consider an assembly scenario where a robot and a human are both placing various pieces on a board. In some cases, pieces may be defective, which the robot can perceive. In this case, the robot has to inform the human about the defective pieces such that the human can replace/discard them. Due to the nature of the environment, which is potentially noisy, the robot may need to use other communication forms than speech signals. For instance, it may instead display a visual clue on a monitor, or move the defective object on a dedicated space as a way to inform the human. The robot therefore has to reason upon the communication actions to perform. In this paper, we discuss some perceptual aspects required for successful communication in such a collaborative setting, and claim that these aspects ought to be taken into account when reasoning about communication activities..

Deliberation is often considered a necessary ingredient for communication in purposeful human-agent interaction [1, 2, 3]. In particular, Sabu *et al* [3] introduce the notion of *deliberative communication* for agent communicators: a reasoning process *before* and *during* communication that addresses questions relating to the effects of a communicative action. These questions include: *What* to communicate, namely, reasoning on the content of the message based on the communicator's knowledge, interlocutor, environment, and context so that the communicator can achieve the communicative goal; *How* to communicate, which includes two dependent levels: inter-how and intra-how, based on the modality used to represent the message (Inter-How) and the expressiveness of the message (Intra-How); and *When* to communicate, namely, identifying the need to communicate (When-Need), and reasoning behind acting (When-Act). As an example, a robot may use reasoning to decide to ask a human operator to carry a given box (*What*), and to ask it now (*When*) using its voice interface (*How*).

HAIC 2025 - First International Workshop on Human-AI Collaborative Systems, editors Michele Braccini, Allegra De Filippo, Michela Milano, Alessandro Saffiotti, Mauro Vallati; October 25, 2025, Bologna, Italy

*Corresponding author.

kiran.mini-sabu@oru.se (K. M. Sabu); Jennifer.Renoux@oru.se (J. Renoux); Hermine.Grosinger@oru.se (H. J. Grosinger); Alessandro.Saffiotti@oru.se (A. Saffiotti)

0009-0005-5111-1629 (K. M. Sabu)

Just performing the action does not guarantee that communication will take place *unless* the action is perceived. In this paper, we consider communication as the action being performed by the communicator *and* being perceived by the interlocutor. In the context of deliberative communication considered here, this means that when deciding to perform communicative actions, the agent may have to reason about the perceiving capabilities of the human *before* and *during* communication to ensure that these actions are perceivable. Interestingly, these capabilities may change dynamically: for example, they may be affected by environmental factors, like noise, or by the cognitive and physical state of the human, like fatigue or impairment. Depending on the capabilities, the human may best perceive actions done via touch, auditory or visual means, and the agent can choose what communicative action to perform accordingly. In our assembly scenario, the robot may decide that the environment is too noisy for a voice message to be perceived, and decide instead to post a visual message on a monitor (*How*) as soon as the human will turn toward it (*When*). After having performed the action, the human may or may not have perceived the action. If the action is not perceived, the message will not be conveyed to the interlocutor. Therefore, the agent also needs to monitor and reason about the uncertainty in the action being perceived by the human after the action is performed. All these reasoning activities require a model of communicative actions that include perception explicitly: this paper reports some initial reflections toward the development of such a model, where, before an artificial agent decides to communicate with a human, the action being perceived must also be considered.

In the field of Human-Robot Interaction (HRI), much attention has been put to the issues of communicating with the human [4] and of perceiving the human [5], but these issues are rarely considered in combination for deliberative communication. Some previous work in HRI [6, 7] considered the perceiving capabilities of a human interlocutor in the context of a robot deliberating for communication as actions. Specifically, Cha *et al.* [6] highlight the importance of perceiving the communicative signal, but reasoning about the signal being perceived by the human was not considered in the planning of communicative actions. Chen *et al.* [7] focus on an assistive shared-control or tele-operation robot setting, and consider the effect of having perceived a communicative action but without considering whether or not the action was actually perceived. Dadvar *et al.* [8] propose a framework for joint communication and motion planning for cobots (Collaborative Robots) that allows robots to communicate to a human during navigation. The sensor model presented accounts for what the human observes and for representing situations where the human many not perfectly understand or observe the robot's communicative actions. However, the robots always have a perfect knowledge of whether the humans perceived and understood the communicative actions, which is a simplifying assumption. In many cases, the robot may not know whether the communicative action has been perceived or understood, and this may result in false-beliefs. Therefore this is an important aspect of deliberative communication which must be considered.

In the rest of this paper, we discuss the ingredients needed to reason about the perception of communication actions in a human-robot collaborative context, and discuss some of the challenges that arise when there is a mismatch between performing a communicative action, perceiving it, and knowing that it has been perceived.

2. Background: Communicative Action

In Speech Acts theory [9] communication is considered as actions, thus embracing the viewpoint that speech actions are performed by agents like other actions. As an agent decides to communicate, it has to perform actions for communication, or *communicative actions*. Building on this, agent communication languages in Multi-Agent Systems (MAS) include communicative actions consisting of a *message* and a *performative* [2]. These approaches typically do not consider the perceptual capabilities of the interlocutor or the fact that the actions are perceived [10].

Existing works have defined communicative actions in various contexts as:

- Hellström *et al.* [11]
  - *"actions performed by an agent, with the intention of increasing another agent's knowledge of the first agent's state of mind."*
- Frijns *et al.* [12]
  - *"actions performed by humans and robots with the aim of coordinating behaviors, reducing uncertainty, and building a common understanding."*
- Sabu *et al.* [3]
  - *"perceivable actions representing a message that can be produced by the communicator and that can produce a change in the mental state of the interlocutor when perceived."*

The definition by Frijns *et al.* [12] does not distinguish between actions and communicative actions. Since we consider the need for deliberation in communication and with agents deciding to communicate, such a distinction is necessary. The definitions by Hellström *et al.* [11] and Frijns *et al* [12] consider the actions being perceived only implicitly. Both definitions focus on the goal which the communication aims to achieve, whereas the definition by Sabu *et al.* [3] focuses on what the action constitutes and what it does by also considering the perceptual aspect.

In this work, we adopt the definition of communicative actions by Sabu *et al.* [3].

3. Communication as Perceived Actions

We now present our preliminary model, that focuses on dyadic communication between a robot agent (*R*) as the communicator, and a human agent (*H*) as the interlocutor. We assume the agents are deliberating and acting in a partially observable environment where actions can have non-deterministic effects. The model below is pre-theoretical, in that we do not propose a full formalization but only present those elements that are the focus of this paper.

3.1. Basic ingredients

Consider a pair of agents $\mathbf{A} = \{H, R\}$, and finite set of states S where $S \subseteq S_O \times S_{E_R} \times S_{E_H}$. S_O are the *ontic* components of states, representing the physical state of the environment and of the agents; S_{E_R} and S_{E_H} are the *epistemic* components of states, representing the mental state of the robot and of the human, respectively. A state $s \in S$ is a triple $\langle s_o, s_{e_R}, s_{e_H} \rangle$. Each agent $i \in \mathbf{A}$ can perform a finite set of actions $A_i = A_{O_i} \cup A_{E_i} \cup \{\alpha_0\}$. A_{O_i} is the set of *ontic* actions that i can perform, i.e., those that result in a change in the ontic state, like the action `MOVE_defective-obj`. A_{E_i} is set of *epistemic* actions that i can perform, i.e., those that result in a change in the epistemic state, like the action `SAY_defective-obj-name`. Note that in cognitive science, the term "epistemic action" is often used as a synonym of information gathering action, that is, an action that only changes the epistemic state of the actor [13, 14]. In our model, we use the terminology from Dynamic Epistemic Logic [15] and postulate that an epistemic action α_i can change the epistemic state of the communicator (s_{e_i}), of the interlocutor (s_{e_j}), or both. Lastly, α_0 is the *null* action that does not produce any change in state, introduced for convenience.

An important feature of actions in our context is that actions may have communicative content. This is obvious for epistemic actions that change the epistemic state of the other agent. However, ontic actions can also have a communicative content and therefore affect the epistemic state: e.g., the previous action `MOVE_defective-obj` may convey the message that the object needs to be replaced. We refer to any action with a communicative content as a *communicative action*. The crucial observation here is that an agreement has been established between the agents on a shared meaning of these actions, which we informally refer to as their *message*. In our model, we assume that such an agreement exists, and that communicative actions are interpreted accordingly by both *H* and *R*. In our example, the robot and the human agree that moving an object to a certain dedicated space means that the object is defective.

3.2. Perceiving actions

We now come to the main point of this paper, that is, the claim that communicative actions must be perceived in order for their effect on the epistemic state of the interlocutor to materialize, and that any deliberative communication should take this fact into account. To model this, we introduce, for any agent $j \in \mathbf{A}$, the following predicate:

- $\texttt{Perceive}_j(\alpha_i, s)$ iff j perceives $\alpha_i \in A_i$ in s performed by i in state s, where $s \in S$, $i \in \mathbf{A}$ and $i \neq j$.

We can use this predicate to give a more precise definition of the robot communicative actions introduced by Sabu *et al.* [3]:

Definition 1. *A communicative action is an action $\alpha_R \in A_R$ representing a message, performed in a state $s \in S$, produces a change in S_{E_H} provided that $\texttt{Perceive}_H(\alpha_R, s)$ is true.*

In the previous example, the robot's executing the ontic action `MOVE_defective-obj` in $s = \langle s_o, s_{e_R}, s_{e_H} \rangle$ could result in $s' = \langle s'_o, s'_{e_R}, s'_{e_H} \rangle$, where s'_o is a new state of the environment and s'_{e_H} is a new epistemic state of the human, provided that the human has perceived the action.[1]

In the previous example, performing α_R also changed the epistemic state s'_{e_R} of the performing agent R. One obvious reason for this change is that R may know that it has performed the action. In this paper, we are more interested in another reason why s'_{e_R} may change: after performing a communicative action, the communicator may have the belief that the action has been perceived, which in turns implies the belief that its epistemic effects have been produced. To model this, we introduce, for any pair of agents $i, j \in \mathbf{A}$ with $i \neq j$, the following predicate:

- $\texttt{Believe}_i\ \texttt{Perceive}_j(\alpha_i, s)$ iff i believes that j perceives $\alpha_i \in A_i$ in s performed by i in s, where $s \in S$.

In general, if a communicative action $\alpha_i \in A_i$ is applicable in a state $s = \langle s_o, s_{e_R}, s_{e_H} \rangle$, then performing α_i in s changes s to a new state $s' = \gamma(s, \alpha_i)$, where γ is a transition function. How the epistemic components of s are affected by the action depends both on $\texttt{Perceive}_j(\alpha_i, s)$ and on $\texttt{Believe}_i\ \texttt{Perceive}_j(\alpha_i, s)$, as summarized in Table 1. Note that how the ontic component s'_o changes only depends on α_i, but is independent on the action's being perceived or not. To illustrate, consider our assembly scenario where the robot has to decide to communicate in state s by either performing the action `MOVE_defective-obj` or `SAY_defective-obj-name`. If the environment is noisy, and the $\texttt{Believe}_R\ \texttt{Perceive}_H(\texttt{SAY\_defective-obj-name}, s)$ is False, the robot can decide to perform the action `MOVE_defective-obj`, provided that $\texttt{Believe}_R\ \texttt{Perceive}_H(\texttt{MOVE\_defective-obj}, s)$ is True. Under these conditions, the action will be perceived, which results in successful communication.

Table 1
The effects of the robot's performing a communicative action depend on the human's perceiving the action, and on the robot's belief that it is perceived.

Case	$\texttt{Perceive}_H(\alpha_R, s)$	$\texttt{Believe}_R\ \texttt{Perceive}_H(\alpha_R, s)$	$\gamma(s, \alpha_i)$
1	True	True	$\langle s'_o, s'_{e_R}, s'_{e_H} \rangle$
2	True	False	$\langle s'_o, s_{e_R}, s'_{e_H} \rangle$
3	False	True	$\langle s'_o, s'_{e_R}, s_{e_H} \rangle$
4	False	False	$\langle s'_o, s_{e_R}, s_{e_H} \rangle$

The γ transition function encodes the effects of actions on states. Some communicative actions may entail a request to the interlocutor to perform an action. For instance, the robot's (epistemic) action `SAY_defective-obj-name` may entail the request to the human to perform the (ontic) action to

[1] A possible variation is one where the message is conveyed by the effect of the action rather than by its mere performance: in that case the `Perceive` predicate should use the resulting state as an argument, but the essence of our discussion would not change.

dispose of the object; the (ontic) action MOVE_defective-obj may entail the same request (assuming the right agreement, as we discussed above); and the (epistemic) action SAY_ask-obj-name may entail the request to perform the (epistemic) action to tell the object name to the robot. In our initial model, we assume that requests to the human will always be complied to: under this assumption, a communicative action α_R done by the robot may result in additional, mediated effects,[2] namely, the effects produced by the action (α_H) requested to (and executed by) the human. We refer to these effects as *after-effects*. Depending on the request contained in α_R, the after-effects can be ontic, epistemic, or null. Also, these effects can be produced in the state s'', resulting from the execution of α_H in state s', or in a later state s''', if the human postpones the execution of α_H: we refer to these as *immediate* and *delayed* after-effects, respectively.

3.3. Misperceiving actions

An important observation regarding the after-effects of a communicative action α_R is that these effects can only materialize if $\texttt{Perceive}_H(\alpha_R, s)$ is true. Modeling this can be important when reasoning about communication.

Consider again the different cases in Table 1. Case 1 is the nominal case, in which the action α_R performed as communication by R results in the action being perceived by H, and R correctly believes that the action has been perceived. Hence, R will expect that the states resulting from the execution of α_R will be affected by both the effects of α_R, and by its after-effects, as we discussed above. In our example, the robot would expect that after SAY_defective-object-name is executed, the object will be disposed of.

In case 4, the action performed by R is not perceived and, correctly, R does not believe that the action has been perceived. Thus, the action performed by the communicator does not result in the change of the mental state of H, nor in the related after-effects. In our example, the object will not be disposed of, and the robot knows that it will not be disposed of: this would enable the robot to, e.g., decide to do the communication again, possibly using a different action.

Cases 2 and 3 result in false beliefs and may lead to problematic situations. These false beliefs may arise due to the partial observability of the actions and epistemic states of other agents. In case 2, although the action performed by R was perceived by H, R has a false belief that it was not perceived. As a result R may decide to perform α_R again, which could result in annoyance on H. In case 3, the action performed by R is not perceived by H, but R has a false belief that is was perceived. As a result, R may move on with execution of other actions that presume that the communication has taken place, or it would wait to observe the effects expected to be produced by H. In our scenario, the robot might have used a SAY_defective-obj-name to communicate a defective object, unaware that the human can not hear that because of the noise level, and then wait forever that the human replaces the object.

Note that the above four cases require different decisions by the robot after α_R has been performed. They may also lead to different decision about which α_R to perform in the first place, depending on the context. These cases of course could not be distinguished if we did not explicitly represent and reason about perception.

4. Next steps

In this work, we discussed the importance of reasoning about communicative actions by a robot collaborating with a human, which — crucially — includes reasoning about whether the communicative action was perceived by the human. Two situations were highlighted in which the robot can have false beliefs where the action may or may not be perceived by the human. As a result, the robot may unnecessarily perform the same action again, or erroneously move on with the execution of other actions, or wait for effects that will not be produced. To address this, the robot has to reason about the uncertainty in the action being perceived by the human. For simplicity, we considered the robot

[2]This relates to the *ramification problem*, which has been extensively discussed in the literature [16].

agent (communicator) performing actions representing a message with the human agent (interlocutor) having to perceive the action. But communication can also be considered as having the message being perceived by the interlocutor instead of the action. Also, while we focused our discussion on human-robot communication, it is worth mentioning that similar considerations (and solutions) could be made for any agent-agent communication, whenever the reliability of communication can be affected by contextual conditions.

The pre-theoretical model presented in this work is not formal. Future work will focus on formalizing the states and actions within the scope of deliberative communication for human-robot collaboration, and use this formalization to address the challenges with respect to the false beliefs of the perceiving of actions. Future models will also address possible delays in perceiving the message. Another challenge that will be addressed is the uncertainty in the human producing the after-effects of communicative actions when the assumption of human complying to a request is relaxed, as in the real-world there is no guarantee that this assumption will always hold.

Acknowledgments

This work was funded by the Swedish Research Council, grants number 2022-04676 and number 2021-05542, and by the European Commission, under the Horizon AI4Europe project, grant agreement 101070000.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] D. E. Appelt, Planning english referring expressions, Artificial Intelligence 26 (1985) 1–33. doi:https://doi.org/10.1016/0004-3702(85)90011-6.

[2] M. J. Wooldridge, An Introduction to Multiagent Systems, 2. ed., repr ed., Wiley, Chichester, 2012.

[3] K. M. Sabu, J. Renoux, A. Saffiotti, Deliberative Communication for Human-Agent Interaction: A Position Paper, in: Proc of the 12th Int Conf on Human-Agent Interaction, ACM, 2024, pp. 11–16. doi:10.1145/3687272.3688299.

[4] A. Bonarini, Communication in human-robot interaction, Current Robotics Reports 1 (2020) 279–285. doi:https://doi.org/10.1007/s43154-020-00026-1.

[5] J. A. Duncan, F. Alambeigi, M. W. Pryor, A survey of multimodal perception methods for human–robot interaction in social environments, ACM Transactions on Human-Robot Interaction 13 (2024) 1–50. doi:https://doi.org/10.1145/3657030.

[6] E. Cha, E. Meschke, T. Fong, M. J. Mataric, A Probabilistic Approach to Human-Robot Communication, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Macau, China, 2019, pp. 6217–6222. doi:10.1109/IROS40897.2019.8968051.

[7] K. Chen, J. Fong, H. Soh, MIRROR: Differentiable Deep Social Projection for Assistive Human-Robot Communication, in: Robotics: Science and Systems XVIII, Robotics: Science and Systems Foundation, 2022. doi:10.15607/RSS.2022.XVIII.020.

[8] M. Dadvar, K. Majd, E. Oikonomou, G. Fainekos, S. Srivastava, Joint communication and motion planning for cobots, in: 2022 International Conference on Robotics and Automation (ICRA), IEEE Press, 2022, p. 4771–4777. doi:10.1109/ICRA46639.2022.9812261.

[9] J. R. Searle, Speech acts: An essay in the philosophy of language, Cambridge university press, 1969. doi:https://doi.org/10.1017/CBO9781139173438.

[10] P. R. Cohen, C. R. Perrault, Elements of a plan-based theory of speech acts, Cognitive Science 3 (1979) 177–212. doi:10.1016/S0364-0213(79)80006-3.

[11] T. Hellström, S. Bensch, Understandable robots - What, Why, and How, Paladyn, Journal of Behavioral Robotics 9 (2018) 110–123. doi:10.1515/pjbr-2018-0009.

[12] H. A. Frijns, O. Schürer, S. T. Koeszegi, Communication Models in Human–Robot Interaction: An Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI), International Journal of Social Robotics 15 (2023) 473–500. doi:10.1007/s12369-021-00785-7.

[13] D. Kirsh, P. Maglio, On distinguishing epistemic from pragmatic action, Cognitive science 18 (1994) 513–549. doi:https://doi.org/10.1016/0364-0213(94)90007-8.

[14] S. Croom, H. Zhou, C. Firestone, Seeing and understanding epistemic actions, Proceedings of the National Academy of Sciences 120 (2023) e2303162120. doi:https://doi.org/10.1073/pnas.2303162120.

[15] T. Bolander, M. B. Andersen, Epistemic planning for single-and multi-agent systems, Journal of Applied Non-Classical Logics 21 (2011) 9–34. doi:https://doi.org/10.3166/jancl.21.9-34.

[16] M. Shanahan, The ramification problem in the event calculus, in: Proceedings of the 16th International Joint Conference on Artifical Intelligence - Volume 1, IJCAI'99, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999, p. 140–146. doi:10.5555/1624218.1624239.