On the Communicative Aspect of Human-Robot Joint Action*

Ross A. Knepper[1]

***Abstract*— Actions performed in the context of a joint activity comprise two aspects: functional and communicative. The functional component achieves the goal of the action, whereas its communicative component, when present, expresses some information to the actor's partners in the joint activity. The interpretation of such communication requires leveraging information that is public to all participants, known as common ground. Humans cannot help but infer some meaning – whether or not it was intended by the actor – and so robots must be cognizant of how their actions will be interpreted in context. In this position paper, we address the questions of why and how robots can deliberately utilize this communicative channel on top of normal functional actions to work more effectively with human partners. We examine various human-robot interaction domains, including social navigation and collaborative assembly.**

I. INTRODUCTION

A current research topic of interest is collaborative behavior for robots working together closely with humans on a joint activity, such as collaborative furniture assembly (Figure 1). A great deal of attention has been paid to what actions to perform [8, 1] and when to perform them [5, 3] in order to complete a cooperative task. Often underappreciated, however, is the implicit communication that occurs as a result of that action *situated in context*. Humans are adept at performing inference as a consequence of observing actions and drawing on common ground – in fact, they instinctively perform this inference, thus reading additional meaning about the intent of an action [2], and many people treat information gleaned in this manner as though it had been stated outright. In this position paper, we argue that to be successful in a joint activity with humans, robots must be cognizant of what messages they will convey from any potential action, and they must select an action that not only achieves the functional goal, but does so in a manner that communicates appropriate information.

This concept is identified by various terms in differing contexts. In robot motion, including reaching [4] and social navigation [12], it has been termed *legibility*. In linguistics, it has been termed *conversational implicature* [7], for which we provide a primer in Sec. III. In each of these cases, the meaning is extracted by leveraging common ground.

In this paper, we contribute:

- a unifying theory of how and why people piggy-back information on top of functional behaviors, and
- example applications of the theory.

*This material is based upon research supported by the Office of Naval Research under Award Number N00014-16-1-2080. We are grateful for this support.

[1]The author is with the Department of Computer Science, Cornell University, Ithaca, NY 14853 rak@cs.cornell.edu

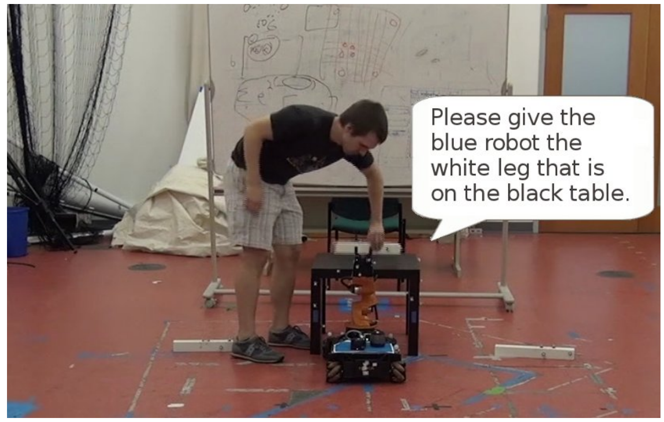

Fig. 1: Robots that collaborate with humans, such as in an assembly task [9], must consider the correctness of both the functional and communicative aspects of their actions.

II. FOUNDATIONS

In the course of a joint activity, a person takes a series of actions $a^1, a^2, \ldots, a^n \in A$, each of which accomplishes both functional and communicative goals to varying degrees. If we consider these actions out of context, each is drawn from a distribution $P(A)$, agreed upon by all parties, describing the frequency with which various actions occur. Even if we restrict the scope of a to actions that accomplish a particular goal, there may be many possible actions to choose from.

Implicit communication is achieved by an actor selecting an action a such that $P(a)$ is small. Note that large $P(a)$ would never be interpreted as communicative because it is expected and thus unnoteworthy. A fuller discussion of these principles can be found in Sec. V. We introduce a context model M as a set of facts that captures information about the environment and the individuals' knowledge. Knowledge that all people know they all share is public knowledge, M_{pub} – this is common ground. Other knowledge is known to be known only by some subset of the participants; q's private knowledge is denoted M^q_{priv}.

Suppose that Alice hopes to convey some information, $m \in M^{Alice}_{priv}$, to Bob without resorting to saying it explicitly. She selects an action a such that $P(a|m, M_{pub}) \gg P(a|M_{pub})$. Since Bob determines a to be an improbable action given what he knows, Bob thus infers that there must be some additional unknown factor m^* that explains a. He then proceeds to infer

$$m^* \leftarrow \operatorname*{argmax}_{m \in M} P(a|m, M_{pub}), \quad (1)$$

for some scope of reasonable contexts M, and thus concludes that $m^* \in M^{Alice}_{priv}$ holds.

We expand on these ideas and provide examples in later sections of this paper, but first we take a brief diversion to introduce a concept from linguistics in order to unify it with notions that have been recently studied in robotics.

III. Implicature Primer

In this section, we give a brief background on *conversational implicature*. We seek to draw parallels between implicature and other methods of coded communication of interest in robotics. Implicature comes from pragmatics, the linguistics subfield that studies the usage of language in context. Basic meaning that is expressed and understood by speech acts is through *entailment* – that is, ideas that logically and unavoidably follow from the words chosen by a speaker.

With implicature, in contrast, the speaker *implicates* (i.e. implies or suggests) an idea without explicitly stating it. It is a frequent phenomenon in English, first described by Grice [7]. Consider this example from Lappin and Fox [11]:

Ann: Do you sell paste?
Bill: I sell rubber cement.
implicature: Bill does not sell paste.

An attribute of conversational implicature in particular is that it is *concealable* – that is, there exists a phrase that, when appended to the sentence, cancels the meaning of the implicature. From the above example, "Bill: I sell rubber cement, which is what you really need for your application."

When it comes to dialog, people have varied and complex motives for implicating meaning rather than entailing it, including politeness, sophistication, succinctness, and social group cohesion. A detailed consideration of these objectives is beyond the scope of this paper.

Grice's *cooperative principle* states, "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" [7]. Indeed, the cooperative principle bears more than a passing similarity to the *pedestrian bargain* of Wolfinger [14], which entreats one both to behave competently and to trust others to behave competently. These principles are both forms of the rational actor assumption.

A vital component of conversational implicature is provided by the four Gricean Maxims, which describe speech that obeys the cooperative principle. The four maxims are

1) Maxim of Quantity: make your contribution as informative as is required (but not more so).
2) Maxim of Quality: Make your contribution one that is true.
3) Maxim of Relation: Be relevant.
4) Maxim of Manner: Be perspicuous. Avoid obscurity or ambiguity; be brief and orderly.

Other maxims have also been proposed, such as "Be polite." Because adherence to the cooperative principle is assumed, utterances can be interpreted in light of these maxims. A speaker can therefore deliberately flout one of the maxims (an improbable action) in order to convey that he is employing implicature. Returning to the previous example, Ann must apply the following inference steps to conclude that Bill does not carry paste.

1) *Contextual premise*: it is mutual, public knowledge that Bill has complete knowledge of the items he sells.
2) *Contextual premise*: there is no contextual relationship linking sales of paste and rubber cement (inclusive or exclusive).
3) Assume Bill follows the cooperative principle and maxims.
4) By (1), Bill can fully resolve Ann's question, and by (3), he will.
5) Only the propositions that Bill does or does not sell paste can completely resolve the question.
6) By (2), there is no way to infer from Bill's answer the proposition that he does sell paste. The cooperative principle forbids obfuscation. Thus, Bill has flouted the maxim of relevance.
7) Therefore, we conclude that Bill does not sell paste.

Conversational implicature is absent when all the maxims are satisfied. One indicates the use of implicature by selecting an action to deliberately flout one of the maxims, or when two maxims conflict and cannot both be satisfied with a single utterance.

An example of the latter occurs in the following exchange:

Mark: Where is the cat?
Sue: The cat is in the hamper or under the bed.
implicature: Sue does not know where the cat is.

Because Sue does not know where the cat is, providing either location alone would violate the maxim of Quality. However, providing both locations conflicts with the maxim of Quantity because the cat is in at most one of the stated locations. Sue chooses to flout the maxim of Quantity.

Let us now consider the purpose for which people choose to employ tools like implicature and legibility when performing actions as part of a team.

IV. Purposes of Communicative Action

Humans are able to express a multitude of ideas "in code", by means other than explicit natural language statements. The primary motivation for encoding communication by embedding it in functional actions is that such communications, being implicit, are often much more efficient. Message categories include expressing intent, coordinating plans, and conveying information. Broadly, these categories all fulfill the role of setting expectations.

Social navigation is the most superficial form of interaction, and yet it is rife with implicit communication. For example, in social navigation, the objective is to avoid collision with co-inhabitants of the space and reach one's destination. Combined, these objectives comprise the navigator's intent. Collision avoidance is only the barest definition of correct navigation – it alone would not be judged as competent behavior by fellow pedestrians [13]. Competence demands that we convey our intended trajectory to nearby observers. We trust in return that they will convey their intent to us. Such intent-expressive actions minimize the global uncertainty about future motions of the agents (humans or robots) in the scene, leading to smooth and stable motion. In the absence of social trust, people begin to behave defensively, and the efficiency of motion drops globally in response.

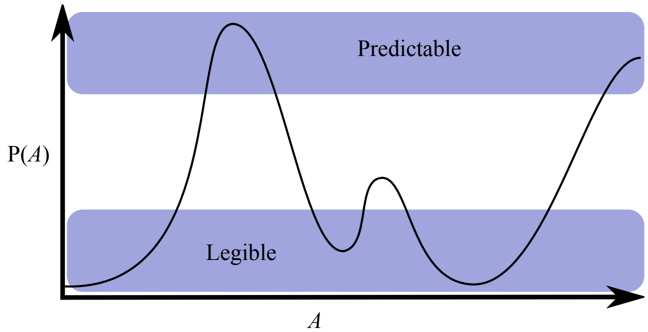

Fig. 2: Among the set of all actions $a \in A$ that accomplish a task, each can be assigned a likelihood of being observed in general, as distinct from in a specific context, $P(a|M)$. The most likely actions are *predictable*, whereas we say that the least predictable actions are *legible*. Since these actions don't ordinarily occur outside of context M, they probably were selected specifically to send a message.

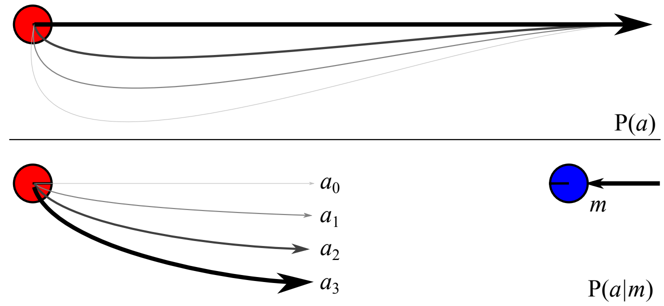

Fig. 3: The red, navigating agent (human or robot) selects an action a. Out of context (top), the red agent (human or robot) is not avoiding an obstacle, and so the probability of expending needless extra energy is low. In the case of an oncoming blue agent (the context m), the likelihood of the oblivious action $P(a_0|m)$ is low due to social norms, despite being low energy. Conversely, the normally-improbable act of spending extra energy becomes probable in this context. An observer who sees only the red agent can infer m from observing a_3.

Coordination among team-mates engaged in a joint activity requires setting expectations of future actions. Consider the simple example of Steve and Cathy assembling furniture together, in which a number of screws must be inserted and tightened. Steve might pick up the screwdriver, which achieves the functional objective of readying Steve to tighten screws. In context, the action also implies that Cathy should gather screws for insertion in order to help. Since Steve is cooperative, Cathy knows that once she begins to insert screws, Steve will fulfill his implicit promise to tighten them.

Beyond forecasting actions, team-mates might also try to convey information about their capabilities. In our model, adjustments to capabilities are an investment whose purpose is to adjust the distribution of future $P(a|M)$ in order to obtain a more appropriate interpretation of action a by the observer. As robots penetrate more deeply into human domains, it will be increasingly necessary for them to carefully calibrate the expectations of humans around them [10]. Robots may appear quite intelligent or human-like in certain respects that may elevate expectations undesirably in other areas. For example, some robots are already imbued with limited natural language processing capabilities, yet humans will likely find discourse with a robot one-dimensional due to limitations in gesture recognition, gaze detection, and context modeling. Properly setting expectations allows robots to avoid disappointing human team-mates.

V. Methods of Communicative Action

In Sec. II, we propose a trigger mechanism for detecting the existence of a communicative component of an action, as well as a method of inferring the meaning of the message. Here, we expand on that discussion.

For functional reasons like energy efficiency, we tend to prefer some actions over others that achieve the same effect, leading to the phenomenon that only a few different actions comprise the majority of observations. The entropy of this distribution over functionally-equivalent actions indicates how much information is conveyed in expectation when one action is chosen. However, some actions, being much less likely, inherently convey less information, and others more. These action categories are given the names *predictable* and *legible*, respectively (Figure 2).

When improbable actions occur, people tend not to interpret them as a fluke. Rather, they assume that some additional (perhaps unknown) information (drawn from common ground) helps to explain the actions and make them probable *in context* (Figure 3). This reasoning is called teleological or goal-directed [2]. Since it is apparently innate in humans, robots can readily exploit it to communicate intent to humans.

In fact, teleological reasoning is a double-edged sword. Robots that act purely functionally and do not consider the way that their actions will be interpreted by humans risk sending essentially random messages encoded through this mechanism. This effect can easily lead to confusion and mistrust on the part of human partners on a team. We can therefore conclude that teleological reasoning will become a compulsory component of autonomous social robots.

Echoing the Gricean Maxims of conversational implicature, we formulate a set of maxims for motion:

1) Maxim of Efficiency: Be parsimonious.
2) Maxim of Motion: Do not collide with objects or obstruct another agent's motion.
3) Maxim of Manner: Be perspicuous and orderly.

These maxims readily come into conflict where multiple agents are present. Much as in the case of implicature, the actor will choose to deliberately flout one of the maxims – typically Efficiency – in order to obey the cooperative principle. It is only by considering the collision-avoidance context that an observer is able to appreciate that by taking an exaggerated trajectory such as a_3 in Figure 3, that the global welfare is improved, as measured by increased energy efficiency and decreased uncertainty.

VI. Other Examples

Let us consider again the joint assembly activity in which Steve and Cathy cooperate to build furniture. Many forms of communicative action arise. One class of actions studied recently by Dragan, Lee, and Srinivasa [4] involves reaching motions. Among parts cluttering a table, Steve has to pick up a particular one. The shape of his reaching trajectory may or may not inform Cathy about Steve's intent. A direct reaching motion is predictable (high probability) and therefore not very communicative. A curved trajectory, in contrast, helps Cathy to identify the target of Steve's reach before he gets there. Much like the social navigation trajectories in Figure 3, more curved trajectories are less probable out of context due to the extra energy they expend. Like in social navigation, an observation of a curved trajectory indicates that the actor is trying to avoid a particular target – in this case because his goal is another target. Cathy's ability to interpret Steve's intent becomes important if Steve's goal is out of reach, in which case his motion becomes a gesture indicating his objective, forming an implicit help request to Cathy.

Teams exchange implicit information in cooperative games as well, when the rules forbid free exchange of information. For example, the bidding conventions of contract bridge allow partners to exchange information about the respective strengths of their hands and arrive at an appropriate contract.

Finally, among married couples, this type of implicit communication eases across all modalities (speech, gesture, gaze, etc.) because spouses develop extremely sensitive models of $P(a|M)$, due to familiarity. Remarkably sophisticated notions can be conveyed by careful action selection in almost any context. We have considerable work before robots can achieve this level of understanding of individual people.

VII. Discussion

Conversational implicature and legibility, though seemingly different domains, are connected by techniques of encoding and decoding meaning using teleological inference. These methods rely heavily on common ground to provide clues about when a message is encoded on an action and what information the message contains. The inference process can be quite complex in real-life situations. Particularly in the case of implicature, many rules must be brought to bear in order to correctly interpret what is being implicated, Goodman and Stuhlmüller [6] show promising early results in modeling a simple form of implicature and performing inference by model inversion.

Interestingly, Knepper et al. [9] show a similar model-inversion approach employed by a robot to generate clear and unambiguous natural language that does not leverage implicature. In this case, the dual components of the robot's action are reversed. The primary role of the action is communicative speech, in the form of a help request. The implicit dual component of the action is functional in that a human hears the request and renders assistance.

In the coming years, modeling of implied meaning, including through implicature and legible motion, will become an increasing focus within robotics. This direction will drive the need for improved modeling of common ground. A major hurdle to performing these inferences on robots in real-world situations is salience – the robot must perform a fairly undirected, brute-force search in order to discover which elements of the context are applicable. Humans, in contrast, seem to learn filters and partially pre-compute functions to expedite real-time inference in ambiguous situations, but these processes are not yet understood.

References

[1] R. Alami, A. Clodic, V. Montreuil, E. A. Sisbot, and R. Chatila. "Task planning for human-robot interaction". In: *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies*. ACM. 2005, pp. 81–85.

[2] G. Csibra and G. Gergely. "'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans". In: *Acta Psychologica* 124.1 (Jan. 2007), pp. 60–78.

[3] M. B. Do and S. Kambhampati. "Planning Graph-based Heuristics for Cost-sensitive Temporal Planning." In: *AIPS*. 2002, pp. 3–12.

[4] A. D. Dragan, K. C. Lee, and S. S. Srinivasa. "Legibility and predictability of robot motion". In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 2013, pp. 301–308.

[5] M. C. Gombolay, R. Wilcox, and J. A. Shah. "Fast Scheduling of Multi-Robot Teams with Temporospatial Constraints." In: *Robotics: Science and Systems*. 2013.

[6] N. D. Goodman and A. Stuhlmüller. "Knowledge and implicature: Modeling language understanding as social cognition". In: *Topics in cognitive science* 5.1 (2013), pp. 173–184.

[7] H. P. Grice. "Logic and conversation". In: *Syntax and Semantics* (1975), pp. 41–58.

[8] B. Hayes and B. Scassellati. "Discovering task constraints through observation and active learning". In: *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE. 2014, pp. 4442–4449.

[9] R. A. Knepper, S. Tellex, A. Li, N. Roy, and D. Rus. "Recovering from Failure by Asking for Help". In: *Autonomous Robots* 39.3 (Oct. 2015), pp. 347–362.

[10] M. Kwon, M. F. Jung, and R. A. Knepper. "Human Expectations of Social Robots". In: *Late Breaking Report at the ACM/IEEE International Conference on Human-Robot Interaction*. Christchurch, New Zealand, Mar. 2016.

[11] S. Lappin and C. Fox. *The handbook of contemporary semantic theory*. John Wiley & Sons, 2015.

[12] C. I. Mavrogiannis and R. A. Knepper. "Towards Socially Competent Navigation of Pedestrian Environments". In: *Robotics: Science and Systems 2016 Workshop on Social Trust in Autonomous Robots*. Ann Arbor, USA, June 2016.

[13] E. A. Sisbot, L. F. Marin-Urias, R. Alami, and T. Siméon. "A Human Aware Mobile Robot Motion Planner". In: *IEEE Transactions on Robotics* 23.5 (2007), pp. 874–883.

[14] N. H. Wolfinger. "Passing moments: Some social dynamics of pedestrian interaction". In: *Journal of Contemporary Ethnography* 24.3 (1995), pp. 323–340.