Implicit Communication in a Joint Action

Ross A. Knepper
Department of Computer Science
Cornell University
Ithaca, NY, USA
rak@cs.cornell.edu

Christoforos I. Mavrogiannis
Sibley School of Mechanical and Aerospace Engineering
Cornell University
Ithaca, NY, USA
cm694@cornell.edu

Julia Proft
Department of Computer Science
Cornell University
Ithaca, NY, USA
jproft@cs.cornell.edu

Claire Liang
Department of Computer Science
Cornell University
Ithaca, NY, USA
cyl48@cornell.edu

ABSTRACT

Robots must be cognizant of how their actions will be interpreted in context. 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. Much of human communication is performed through this implicit mechanism, and humans cannot help but infer some meaning – whether or not it was intended by the actor – from most actions. We present a framework for robots to utilize this communicative channel on top of normal functional actions to work more effectively with human partners. We consider the role of the actor and the observer, both individually and jointly, in implicit communication, as well as the effects of timing. We also show how the framework maps onto various modes of action, including natural language and motion. We consider these modes of action in various human-robot interaction domains, including social navigation and collaborative assembly.

1. INTRODUCTION

An important domain for human-robot interaction involves collaboration on a joint activity, such as collaborative furniture assembly (Figure 1). A great deal of attention has been paid to what actions to perform [1, 16, 17, 27] and when to perform them [9, 13, 35] in order to complete a cooperative task. Often underappreciated, however, is the implicit communication that occurs as a result of an *action situated in context*. We call behaviors that implicitly communicate information *communicative actions*. Humans are adept at drawing inference from observed actions and common ground – in fact, they instinctively perform this inference, thus reading additional meaning about the intent of an action [8], and many people treat information gleaned implicitly through inference as though it had been stated outright. We argue that to be successful in a joint activity with humans, robots must be cognizant of implicit communication because humans will inevitably use it and expect robots to comprehend its meaning. We further argue that if a robot fails to attend to a human's interpretation of its own actions through the implicit communication mechanism, then people will perceive the robot's purely functional actions as sending random implicit signals, sowing confusion.

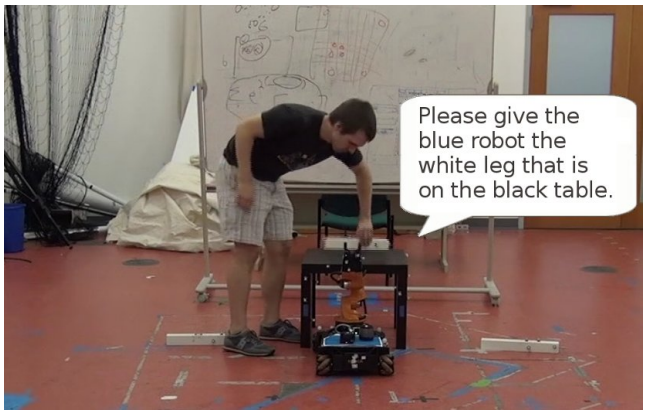


Figure 1: Robots that collaborate with humans, such as in an assembly task [22], must consider the correctness of both the functional and communicative aspects of their actions.

Implicit communication is identified by various terms in differing contexts. In robot motion, including reaching [10] and social navigation [29], it has been termed *legibility*. In linguistics, it has been termed *conversational implicature* [15], for which we provide a primer in Sec. 5.1.1. In natural language generation for HRI, it has been called *inverse seman-*

HRI '17, March 06 - 09, 2017, Vienna, Austria

ISBN 978-1-4503-4336-7/17/03…$15.00

DOI: http://dx.doi.org/10.1145/2909824.3020226

tics [22]. In each of these cases, the meaning is extracted by leveraging common ground. The goal of this paper is to unify these separate works by explicating a common mathematical framework that underlies all of them.

Extending an earlier workshop paper [20], we contribute:

- a unifying mathematical framework describing how and why people implicitly communicate information on top of functional behaviors,
- formal expressions for encoding and decoding communicative actions, and
- collected example applications to illustrate the theory.

2. WHY IMPLICITLY COMMUNICATE?

Humans are able to express a multitude of ideas "in code", by means other than explicit natural language statements. Motivations for implicit communication include efficiency, tact, group cohesion, and social bonding. In this section, we give examples of several categories of implicit communication. Message categories include expressing intent, coordinating plans, and conveying information. Broadly, these categories all fulfill the role of setting expectations, and we consider each separately.

Social navigation is among the most superficial forms of interaction, yet it is rife with implicit communication. 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 without intent expression is only the barest definition of correct navigation – it alone would not be judged as competent behavior by fellow pedestrians [31]. 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. We borrow from Barbalet [2] the definition of trust as "the confidence that another's actions will correspond with one's expectations." In the absence of social trust, people begin to behave defensively, and the efficiency of motion drops globally in response.

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. Human interactional expectations are broadly governed by a common set of human functional and social capabilities, whereas humans are largely uninformed about a robot's true capabilities. Therefore, robots will likely find themselves being judged according to the wrong standards. Although humans show patience for robots that fail under the right conditions, a robot that seldom works as expected will likely not remain in use, even if the failure is one of expectations rather than capabilities. Properly setting expectations allows human team-mates to avoid being disappointed by robots [5, 23, 25].

3. FRAMEWORK

In this section, we describe a framework for implicit communication, modeled as a single-shot act.

3.1 Definitions

In coordinated activities, Clark [6] distinguishes among several related concepts. A *joint activity* engages a group of two or more agents in acting together toward a common goal. Examples include a marriage ceremony, a classroom lecture, and a football game. Within the context of a joint activity, participants perform *joint actions*, which continuously unfold over some period of time. A specialization is the *joint act*, which is a one-shot joint action. For example, in the joint activity of playing golf, yelling "fore!" is a joint act by which the speaker warns any listeners of a wayward flying ball (their avoidance response, in contrast, is an individual act, performed without consideration of how it will affect the group). The fact of an act being joint or individual is purely a matter of the mental state of the involved agent(s).

Participation in a joint action may be asymmetric – for example, speech is a joint action involving a speaker and listener. Note that the listener actively participates by comprehending and back-channeling (nodding, saying "uh-huh", etc.). *Knowledge* comprises information believed by an agent to be true and is collected into a set of *facts*, each with associated confidence. Compulsory asymmetry occurs in a joint act or action when one individual, the *actor*, shares knowledge with one or more *observers*. Thus, an important aspect of the joint action is to communicate information. Frequently, an actor embeds information implicitly in an otherwise purely functional action as part of the joint activity to perform *implicit communication*.

Any communicative action will be perceived by an observer with a certain level of *surprisal*, which is an encoding of how probable the observer believes the action to be given the context. As Hohwy [19] states, surprisal is a declining function of probability: the higher an observer's surprisal, the more improbable the observer believes the action to be in the given context; the lower an observer's surprisal, the more probable the observer believes the action to be in the given context. Common-sense knowledge and a shared understanding of the context allows an actor to gauge how surprising her action will be to an observer, which in turn shapes her choice of action depending on the information she would like to convey. In the remainder of this section, we show that greater surprisal corresponds with a more strongly-conveyed message (i.e. the action is more meaningful).

3.2 Foundations

The interplay of two sets is at the core of the framework. $\mathcal{A}$ comprises all possible actions, whereas $\mathcal{M}$ is composed of all possible facts about the world.

In the course of a joint activity, an agent performs a series of actions (including single-shot acts) $a^1, a^2, \ldots, a^n \in \mathcal{A}$. Each action accomplishes both functional and communicative goals to varying degrees. Let $A_f \subseteq \mathcal{A}$ be the set of (possibly many) different ways of accomplishing the functional goal of the action. Thus, A_f can be thought of as a subgoal of the shared goal of the joint activity.

An agent $\mathcal{Q}$ performs actions in a context $M^{\mathcal{Q}}$ comprising a set of facts $m_1, m_2, \cdots \in M^{\mathcal{Q}} \subset \mathcal{M}$ that capture information about the individuals' knowledge. Only by leveraging this context can implicit communication occur. $M^{\mathcal{Q}}$ ex-

presses $\mathcal{Q}$'s beliefs about the world, including the state history of all agents in the joint activity, the observable scene, properties of objects within it, and common-sense knowledge. An individual fact $m \in M^{\mathcal{Q}}$ can have an associated confidence, thus allowing facts in $M^{\mathcal{Q}}$ to be added, removed, or changed following the observation of an action.

$M^{\mathcal{Q}}$ is divided into several components. Knowledge that all participants in an interaction know they all share is public knowledge, M_{pub}, also called common ground. Other knowledge is not known to be public; agent $\mathcal{Q}$'s private knowledge is denoted $M^{\mathcal{Q}}_{priv}$. $\mathcal{Q}$ may be aware that a subset of the other agents know fact $m \in M^{\mathcal{Q}}_{priv}$. It is even possible that every agent in a joint activity privately knows m. In both cases, $m \notin M_{pub}$ unless all agents are all aware that m is shared by all. $\mathcal{Q}$'s total knowledge $M^{\mathcal{Q}}$ is equal to $M_{pub} \cup M^{\mathcal{Q}}_{priv}$.

Finally, the distribution $\mathrm{P}(a|M)$ describes the likelihood that a specific agent may next perform action a in the specific context M. Even if we restrict the scope of a to actions that accomplish a particular goal, there may be a set of possible actions ($A_f \subseteq \mathcal{A}$) to choose among. Some of these actions will be preferred over others for reasons of efficiency, simplicity, or custom.

Posit that the following *common understandings* are agreed upon by all participants in the joint activity:

- the set of alternative actions A_f that would accomplish a functional goal
- the common ground context model M_{pub}
- the action distribution $\mathrm{P}(a|M)$ (for plausible $M \subset \mathcal{M}$)

3.3 Implicit Communication Criteria

The goal of agent $\mathcal{Q}$ is to perform an action $\hat{a} \in A_f$ that satisfies functional goals while also communicating fact $\hat{m} \in M^{\mathcal{Q}}_{priv}$. However, it is not always possible to communicate an arbitrary fact $\hat{m}$ implicitly, nor is it always possible to communicate implicitly via an action $\hat{a}$.

The key idea is for the actor $\mathcal{Q}$ and observer $\mathcal{R}$ to leverage the common understandings in order to achieve implicit communication. $\mathcal{Q}$ selects an action that is surprising to $\mathcal{R}$, i.e. perceived by $\mathcal{R}$ as improbable in the given context. However, $\mathcal{R}$ does not treat the improbable $\hat{a}$ as a fluke – rather, it triggers her to seek an explanation in the form of a previously-unknown fact $\hat{m}$ that resolves the surprise. For $\mathcal{R}$ to correctly interpret $\mathcal{Q}$'s intended meaning, we propose that action $\hat{a}$ and fact $\hat{m}$ must meet four *implicit communication criteria*:

1. $\exists \hat{a}, a' \in A_f : \hat{a} \neq a'$
2. $\mathrm{P}(\hat{a}|M_{pub}) < \mathrm{P}(a'|M_{pub}) - \varepsilon$
3. $\mathrm{P}(\hat{a}|M_{pub}) < \mathrm{P}(\hat{a}|\hat{m}, M_{pub}) - \varepsilon$
4. $\forall m \in \mathcal{M} \backslash M_{pub} \cup \{\hat{m}\} : \mathrm{P}(m|\hat{a}, M_{pub}) < \mathrm{P}(\hat{m}|\hat{a}, M_{pub}) - \varepsilon$

The ε term incorporates variation caused by personal preference and noise. The strength of a given implicit communication is measured as the largest possible ε satisfying the criteria above. Criteria 1–2 govern the actor's generation of implicit communication, whereas criteria 3–4 govern the observer's ability to correctly interpret the intended meaning. We speak of the fact $\hat{m}$ as the *meaning* of the action because it *explains* $\mathcal{Q}$'s choice of action. We next provide additional insight into each of the criteria.

Criterion 1 requires that there must be at least two feasible, distinct actions that accomplish the functional goal, but preferably there are many more. An example of A_f that violates this criterion is placing a telephone call. Neglecting timing and caller ID, there is only one way to make somebody's telephone ring, leaving no room for a surprising choice of action.

Criterion 2 triggers the observer to search for an explanation of why the actor chose action $\hat{a}$ over the more obvious choice, a'. This criterion fails in situations where there does not exist an action $\hat{a}$ that is a priori substantially less probable than others. An example situation that violates it is one's first time visiting a clown convention, where normally-improbable actions are expected and hence unsurprising.

Criterion 3 requires that the fact $\hat{m}$ will be easy for the observer to verify as an explanation of $\hat{a}$. That is, $\hat{a}$ is unsurprising when $\hat{m}$ is known. A well-known historical violation of this criterion was John Hinckley, Jr.'s attempted assassination of President Ronald Reagan in order to gain the favor of actress Jodie Foster – it is unclear how shooting the president is intended to convey infatuation.

Criterion 4 states that no other inferred meaning m is equally or more likely than the intended explanation $\hat{m}$. There are many example violations of this criterion in the form of hand gestures that take different meanings across cultures and geographies. One case in point is a gesture that variously signifies a Satanic association, infidelity, and a college football team in Texas. All three forms have famously been used by politicians. Only by understanding each individual actor's M_{pub} at the time he made the gesture can we disambiguate among the three meanings.

3.4 Understanding and Generation

Suppose that an agent $\mathcal{Q}$ hopes to convey some information, $\hat{m} \in M^{\mathcal{Q}}_{priv}$, to agent $\mathcal{R}$ without resorting to disclosing it explicitly. $\mathcal{Q}$ selects an action $\hat{a}$ consistent with the implicit communication criteria, and $\mathcal{R}$ determines $\hat{a}$ to be an improbable action given what he knows. $\mathcal{R}$, believing $\mathcal{Q}$ to be rational, hypothesizes that there must be some unknown factor $\hat{m}$ that explains seeing $\mathcal{Q}$ perform $\hat{a}$. $\mathcal{R}$ thus searches over a set of plausible facts M and chooses $\hat{m}$ to be the fact with the highest posterior probability given $\hat{a}$ and M_{pub}. Maximizing this probability minimizes the surprisal that resulted from $\mathcal{Q}$ performing $\hat{a}$, which in turn causes $\hat{a}$ to become increasingly stronger evidence for $\mathcal{R}$'s hypothesis [19]. Hence, upon seeing $\hat{a}$, $\mathcal{R}$ proceeds to infer

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

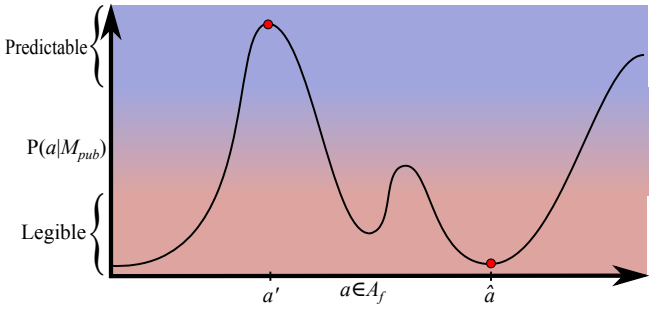
and thus $\mathcal{R}$ concludes that $\hat{m} \in M^{\mathcal{Q}}_{priv}$, i.e. $\mathcal{Q}$ believes $\hat{m}$ to be true. Using Bayes' rule, we can re-express (1) as

$$\begin{aligned} \hat{m} \leftarrow & \operatorname*{argmax}_{m \in M} \frac{\mathrm{P}(\hat{a}|m, M_{pub})\mathrm{P}(m|M_{pub})}{\mathrm{P}(\hat{a}|M_{pub})} \\ = & \operatorname*{argmax}_{m \in M} \mathrm{P}(\hat{a}|m, M_{pub})\mathrm{P}(m|M_{pub}). \end{aligned}$$

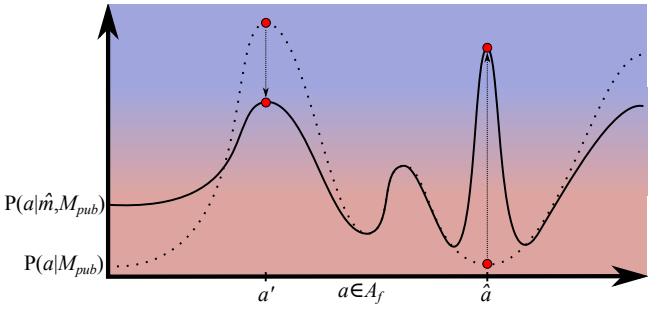
Note that the prior $\mathrm{P}(m|M_{pub})$ serves to prevent "conspiracy theories" that would otherwise result when noise gets mistakenly interpreted as signal. That is, the fact being communicated must have a reasonably likely prior probability. For example, if Bob looks up at the night sky and sees a star twinkling, he is unlikely to attribute it to a UFO, given that the prior probability of discovering intelligent extraterrestrial life is small and that there is a more plausible explanation rooted in turbulence of the atmosphere.

Next, we turn to the generation problem. The structure of the generation problem is identical to understanding, except that we now search over actions instead of facts,

$$\hat{a} \leftarrow \operatorname*{argmax}_{a \in A_f} \mathrm{P}(\hat{m}|a, M_{pub}). \tag{2}$$



(a) A likely action such as a' is termed *predictable*, whereas we say that an unlikely action $\hat{a}$ is *legible*. See Section 5.2.1 for a full discussion of predictability and legibility. Since $\hat{a}$ is rarely observed in context M_{pub}, the observer infers that it probably was selected specifically to send a message.



(b) By performing legible action $\hat{a}$, an agent implicates the new fact $\hat{m}$ because knowledge of that fact changes the shape of the distribution, causing $\hat{a}$ to become a predictable action.

Figure 2: These plots illustrate the inference mechanism described in Section 3.4 and its effect on $\mathrm{P}(a|M_{pub})$. Among the set of actions $a \in A_f$ that accomplish a task, each can be assigned a likelihood of being observed in context. Actions with high likelihoods of $P(a|Mpub)$ are deemed *predictable*, low ones *legible*.

Applying Bayes' rule again, we can re-express (2) as

$$\hat{a} \leftarrow \underset{a \in A_f}{\text{argmax}} \frac{\mathrm{P}(a|\hat{m}, M_{pub})\mathrm{P}(\hat{m}|M_{pub})}{\mathrm{P}(a|M_{pub})}$$
$$= \underset{a \in A_f}{\text{argmax}} \frac{\mathrm{P}(a|\hat{m}, M_{pub})}{\mathrm{P}(a|M_{pub})}.$$

The resulting expression selects the action for which contributing $\hat{m}$ to the common ground boosts $\mathrm{P}(a|M_{pub})$ by the greatest amount. See Figure 2 for an illustration.

We expand on these ideas and provide examples in Section 5, but first we broaden our discussion to include implicit communication occurring over time and in service of joint goals.

4. ACHIEVING JOINT GOALS

In a joint activity, rational agents interact with each other and make decisions towards achieving joint goals. These goals could range from completing a collaborative assembly task to smoothly avoiding each other while navigating in a hallway. Relying only on implicit communication to achieve joint goals requires the establishment and reinforcement of trust. Implicit communication leverages trust to influence the observer's belief and converge to a consensus that is beneficial for the accomplishment of a joint goal. In this section we state our model for trust and propose an index for monitoring its evolution in order to inform decision making.

4.1 Trust

Ordinarily, participants in a joint activity act *rationally* and cooperate to achieve shared goals [18]. This policy forbids deception and supports the assumption that the common understandings (Section 3.2) are shared by all participants. Given the great diversity of knowledge and experience among people, however, this assumption is perhaps too strong to apply universally.

In particular, during interactions with strangers, we may be unfamiliar with one another's judgments regarding A_f, M_{pub}, and $\mathrm{P}(a|M_{pub})$. If we define trust as confidence in another agent's future actions [2], then it is natural for one agent to restrict their trust of another based on the limits of common understandings among the individuals, even when all agents behave rationally.

Another obstacle to trust is discrepant beliefs about facts. We allow facts about the beliefs of others to enter M_{pub}. Thus, it can simultaneously be part of the common ground that $\mathcal{G}$ believes $m^{\mathcal{G}}$ and that $\mathcal{H}$ believes $m^{\mathcal{H}}$, even if $m^{\mathcal{G}}$ and $m^{\mathcal{H}}$ conflict. $\mathcal{G}$ and $\mathcal{H}$ are then free to leverage either of these facts in the generation and understanding of implicit communications between them. Epistemic logic [11] provides tools for representing and analyzing such scenarios. Each conflicting fact introduces additional uncertainty into the communication process because the observer must infer which fact the actor premised the communication upon. Thus, trust degrades with the number of discrepancies among beliefs within a joint activity. Beyond some limit, implicit communication becomes impossible.

4.2 Consensus

In a joint activity, agents take actions with functional effects (which contribute to reaching the joint goal) but also with communicative effects. One category of communication, conveying intentions, serves to convey a preference or desire regarding a joint strategy S for accomplishing the goal. The joint strategy can be thought of as the sequence of subgoals of the joint activity, $A_f^0, A_f^1, \ldots, A_f^n$, and is drawn from the set of all possible strategies $\mathcal{S}$.

A consensus for each subgoal in the joint strategy may unfold gradually or abruptly during the course of the joint activity. As the agents act, the public knowledge M_{pub} is updated along with the agents' beliefs regarding the emerging strategy $\mathrm{P}(S|M_{pub})$. Under the assumption of rationality, as formulated in our trust model (Section 4.1), a group of competent agents taking actions bearing implicit communication signals will be able to achieve consensus over the joint strategy S. This essentially means that $\mathrm{P}(S|M_{pub})$ (which we assume is shared by all agents) will converge to a distribution that clearly indicates the emerging joint strategy. The entropy of this distribution is a measure of that convergence.

4.3 Receptivity

In many joint activities, time and timing are critical attributes of an action. Timing itself often conveys mean-

ing, which we therefore consider as an attribute of an action $\hat{a} \in A_f$. Another important aspect of timing is its role in choosing whether (and when) to implicitly communicate. Participants in a joint activity are not equally receptive at all times to certain forms of implicit communication, particularly with regard to consensus over the joint strategy.

When participants in a joint activity lack consensus about a joint strategy, they cannot coordinate effectively to achieve shared goals. Rational agents therefore strive to reach consensus as early in a joint activity as possible in order to maximize coordination efficiency. Consequently, the bulk of implicit communication for consensus should occur towards the beginning of the joint activity. As a joint strategy S^* emerges and consensus is reached, the agents might favor more predictable, less communicative actions, or they might utilize the implicit communication channel for other purposes. More generally, the implicit consensus formation aspect of joint actions may wax and wane according to the group need. Consequently, there arises the need for monitoring (1) the state of consensus $\mathrm{P}(S|M_{pub})$ and also (2) how receptive the group of agents is to the communicative signals being transmitted.

We formalize this monitoring process by introducing a *Receptivity* score, as

$$\textit{Receptivity} = -\sum_{S \in \mathcal{S}} \mathrm{P}(S|M_{pub}) \log(\mathrm{P}(S|M_{pub})) \qquad (3)$$

which is the information entropy of the distribution over joint strategies, given the common ground, $\mathrm{P}(S|M_{pub})$. Recall that common ground includes the action history within a joint activity. Intuitively, receptivity measures the willingness of individuals in a group to update their beliefs about the joint strategy, inversely proportionate with clarity. Since M_{pub} is sequentially updated over time, receptivity reflects the way the agents incorporate observed communicative signals into their own actions. The lower a receptivity score gets, the closer the agents are to a consensus over a joint strategy S^*. To avoid second-guessing a settled joint strategy, an observer suppresses strategy changes of a larger magnitude than the current receptivity level.

A consequence of a decline in receptivity is that agents can be less expressive when it drops, since other agents will likely ignore the inputs. In a scene with engaged competent agents, receptivity is expected to decrease rapidly, signifying a consensus in the joint activity. This decrease will influence the balance between the functional and communicative aspect of actions taken, shifting the focus of decision making towards the functional component. Beyond some threshold drop in receptivity, agents have become sufficiently certain about the consensus strategy S^* that they may even ignore their partners using *civil inattention* [21] to reinforce the previously agreed strategy. This behavior involves physically looking away, "so as to express that [one] does not constitute a target of special curiosity or design" [12]. At this point, only a major modification in the strategy will penetrate an agent's civil inattention.

5. CASE STUDIES

In lieu of generating new experimental results, which would apply to a single domain and communication modality, we present examples of how implicit communication has been modeled and enforced by several communities in various collaborative contexts and discuss how their frameworks align with our unifying framework for implicit communication.

5.1 Implicit Communication through Natural Language

Speech acts are among the richest functional actions in which to embed implicit communication.

5.1.1 Implicature

In this section, we give a brief background on *conversational implicature*. We seek to draw parallels between implicature and other methods of implicit 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 a speech act is achieved by *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 [15]. Consider this example from Lappin and Fox [24]:

> Ann: Do you sell paste?
> Bill: I sell rubber cement. ($\hat{a}$)
> *implicature: Bill does not sell paste. ($\hat{m}$)*

A test for conversational implicature in particular is whether it is *cancelable* – that is, does there exist some phrase that, when appended to the sentence, cancels the meaning of the implicature? From the above example, a phrase that cancels Bill's implicature is "I sell rubber cement, which is what you really need for your application." An implicature, once canceled, implicitly communicates nothing. The added phrase explains the initial phrase, thus increasing $\mathrm{P}(a|M_{pub})$ and violating implicit communication criterion 2.

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" [15]. Indeed, the cooperative principle bears more than a passing similarity to the *pedestrian bargain* of Wolfinger [36], which entreats one to behave competently and also to trust others to behave competently. These principles are both forms of the rational actor assumption [18].

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, $\hat{a}$) 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.

(a) *Contextual premise*: it is mutual, public knowledge that Bill has complete knowledge of the items he sells.
(b) *Contextual premise*: there is no contextual relationship linking sales of paste and rubber cement (inclusive or exclusive).
(c) Assume Bill follows the cooperative principle and maxims.
(d) By (a), Bill can fully resolve Ann's question, and by (c), he will.
(e) Only the propositions that Bill does or does not sell paste can completely resolve the question.
(f) By (b), 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.
(g) Therefore, we conclude that Bill does not sell paste.

Lines (d)–(g) comprise the narrowing down and resolution of the search for meaning in Equation (1).

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 – in our example, Bill flouts the maxim of Relation.

Sometimes, two maxims conflict and cannot both be satisfied with a single utterance, in which case flouting one or the other maxim is forced. 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. ($\hat{a}$)
> *implicature: Sue does not know where the cat is.* ($\hat{m}$)

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. Flouting the maxim of Quality would violate implicit communication criterion 2 because either location alone is plausible. Thus, Sue chooses to flout the maxim of Quantity in order to trigger Mark to search for an explanation.

5.1.2 Inverse Semantics

Though more direct than conversational implicature, the simpler speech act of entailment is fundamentally described by the same mathematics. Knepper et al. [22] present the *inverse semantics* framework for robots generating natural language help requests. Like most robot speech systems, the framework strives for extremely literal communication. However, it faces a problem of finding pithy, unambiguous means of communicating its needs in an automated assembly scene cluttered with parts that lack unique names. Since words are complex and imperfect containers for meaning, the careful selection of clear language to achieve entailment follows the same rules of generation as described in Section 3.4.

The core of inverse semantics is a forward semantics mechanism for understanding natural language, the Generalized Grounding Graph (G^3) [32]. This structure takes in natural language expressions λ as inputs and returns their meanings or groundings γ as outputs.

The inverse semantics framework inverts G^3 to perform generation by searching over the space of possible English sentences, sorted from shortest to longest, and inputting each to G^3. Inverse semantics compares the output of G^3 with the target grounding needed by the help request. The search halts with the first sentence that attains over a threshold confidence match between the two groundings. The expression given for generation,

$$\underset{\lambda}{\operatorname{argmax}}\, \mathrm{P}(\gamma|\lambda, \phi, M), \tag{4}$$

strongly resembles our own framework's Equation (2). Here, ϕ is a correspondence variable used to indicate the semantic likelihood of a match between λ and γ. Like our model, M symbolizes the context model in which the meaning is interpreted.

5.2 Communicative Motion

Besides natural language usage, the robotics community has studied other types of actions. An especially expressive action class for implicit communication is motion.

5.2.1 Legibility

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 [10] 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 $\mathrm{P}(a|M_{pub})$) and therefore not communicative. A curved trajectory, in contrast, helps Cathy to identify the target of Steve's reach before he gets there.

In general, assume that an actor $\mathcal{Q}$ is aiming at reaching a goal $G^{\mathcal{Q}}$ from a set of goals $\mathcal{G}$ in front of an observer $\mathcal{R}$. The agents share a model $\mathrm{P}(G|\xi)$ that probabilistically attributes a goal $G \in \mathcal{G}$ to an observed trajectory ξ. The actor can leverage this knowledge to design his trajectory in a way that indicates his intended goal to the observer. Following the insights of Csibra and Gergely [8] regarding the tendency of humans to interpret observed actions as goal-directed (teleological reasoning), Dragan, Lee, and Srinivasa [10] introduced the *Legibility* score to quantify the intent-expressiveness of a trajectory ξ with respect to a goal $G^{\mathcal{Q}}$:

$$\textit{Legibility}(\xi) = \frac{\int_0^T \mathrm{P}(G^{\mathcal{Q}}|\xi_{0\to t}) f(t) dt}{\int_0^T f(t) dt} \tag{5}$$

where T is the duration of the trajectory and $f(t)$ is a function that weights partial trajectories $\xi_{0\to t}$ higher in the beginning and lower later. It should be noted that $f(t)$ is a proxy for the role of the observer in reducing her receptivity (see Section 4.3) as $\mathcal{Q}$'s intended goal $G^{\mathcal{Q}}$ becomes more certain to her. The model $\mathrm{P}(G|\xi)$ scores goals higher if they can be achieved efficiently (with a low energy trajectory ξ) and scores goals lower if they require higher energy.

The legibility score is essentially a weighted sum of the probabilities that the observers assign to the actor's intended goal $G^{\mathcal{Q}}$ throughout the whole trajectory ξ. Trajectories of higher legibility tend to be more curved towards the intended goal $G^{\mathcal{Q}}$, biasing the observers towards predicting the actor's actual goal, while biasing them against predicting other goals. Note that a more curved trajectory is less probable out of context due to the extra energy it expends. As a result, it might be perceived as surprising. This surprise would trigger a search for an explanation, which, in the perceived context, would lead to the conclusion that the actor $\mathcal{Q}$ is aiming at reaching the goal $G^{\mathcal{Q}}$.

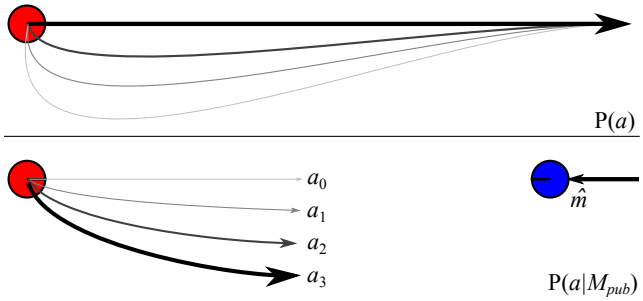


Figure 3: The red, navigating agent (human or robot) selects an action $\hat{a}$. Out of context (top), the red agent (human or robot) is not avoiding an obstacle, and so the probability of expending needless extra energy is low. In the case of an oncoming blue agent ($\hat{m}$), the likelihood of the oblivious action $\mathbf{P}(a_0|M_{pub})$ 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's motion can infer $\hat{m}$ from observing a_3.

5.2.2 *Dynamic Legibility*

Consider now the case of a dynamic environment, where the agents are not explicitly collaborating but since the decisions they make are coupled, it is beneficial for everyone to mutually agree on a joint strategy. Assuming again no explicit communication, the only way agents are able to agree on a strategy is to encode their understanding and preferences into their actions.

Social navigation constitutes a representative example of this class of scenarios. Although humans might not often realize that navigation in crowded environments is a collaborative activity, according to sociology studies [36], it is established on implicit cooperation. Pedestrians follow and reinforce the *pedestrian bargain*, a social convention comprising two foundational rules: (1) "pedestrians must behave like competent pedestrians" and (2) "pedestrians must trust that co-present others behave like competent pedestrians". Since the pedestrian bargain serves as a cooperative principle for social navigation, we may formulate a set of maxims for motion that echo the Gricean Maxims of conversational implicature,

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 the maxim of 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, the global welfare is improved, as measured by increased energy efficiency and decreased uncertainty.

Enforcing the pedestrian bargain leads to a consensus over a mutually beneficial joint strategy that allows everyone to comfortably reach their destinations. The agents continuously monitor the progress toward consensus and adjust their decision-making accordingly. Once consensus appears to have been reached, receptivity drops to zero as pedestrians initiate *civil inattention* [12, 21]. Following this mode switch, agents look away from one another as a signal that they have stopped actively avoiding each other and will instead follow their previous planned collision-free path.

Mavrogiannis and Knepper [28, 29] present a recent navigation framework that reproduces the implicit communication of social navigation. They model consensus S as a topological joint strategy using braids [3], and they develop a game-theoretic decision making policy corresponding to a utility function that compromises between efficiency and implicit communication. This utility function is defined as:

$$u(a) = \lambda E(a) - (1-\lambda)H(a) \tag{6}$$

where λ is a weighting factor, E represents the efficiency of an action a and H is the entropy of the agents' belief regarding the emerging strategy S:

$$H(a) = -\sum_{S\in\mathcal{S}} \mathrm{P}(S|Z^+, M)\log \mathrm{P}(S|Z^+, M) \tag{7}$$

with Z^+ comprising Z (the state history of all agents so far), augmented with the action in consideration a and M being the context.

Picking actions that maximize the utility leads to a quick decrease in the uncertainty regarding the global joint strategy (i.e., how each agent will avoid each other), while ensuring progress towards the agent's destination. The selection of actions with a strong communicative component, especially in the beginning, was shown to reduce the likelihood of livelocks or deadlocks.

6. OTHER EXAMPLES

Teams exchange implicit information in cooperative games 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 over time across all modalities (speech, gesture, gaze, etc.) because spouses develop extremely sensitive models of $\mathrm{P}(a|M_{pub})$, due to familiarity. Remarkably sophisticated notions can be conveyed between spouses by careful action selection in almost any context. We have considerable work remaining before robots can achieve a similar level of understanding of people.

6.1 Tact

Implicit communication is also the primary tool of tactful communication, as it alleviates the risk of awkwardness due to misunderstandings about what facts the observer already knows. Reflecting on the implicit communication criteria given in Section 3.3, an attempted implicit communication of a fact that the observer already knows does not even seem like implicit communication – it would come across as a predictable, functional action. In this case, criterion 3 is clearly violated because $\hat{m} \in M_{pub}$, and criterion 2 is probably also violated because $\hat{a}$ would seem likely.

To offer a concrete example of how speakers leverage implicit communication to achieve tact, consider a married couple discussing dinner plans:

> Jack: Remember, my friend Irving is coming for dinner.
> *implicatures: Irving is vegetarian; Irving needs*

to be served a vegetarian meal.
Kate: Let's make my mother's lasagna recipe.
implicatures: Kate knows that Irving is vegetarian; Kate's mother's lasagna recipe is vegetarian; the recipe satisfies Irving's need for a vegetarian meal.

Observe that this exchange can be read at two levels. If both parties are oblivious to the implicature because the sentences are judged predictable, then it is a simple, matter-of-fact dialog.

The statements can also be read as implicature. In both cases, the implicated statements are things that the listener should have already known. Only in the context of the couple's normal conversation can we judge how unusual it is for Jack to remind Kate about a guest (a fact she may be unlikely to forget), or for Kate to make her mother's lasagna recipe.

Only if these events are atypical can they truly be regarded as implicit communication. However, they also serve a tactful reminder function, in case Kate forgot about the guest or Jack forgot that Kate's mother's lasagna is vegetarian. A failing memory may therefore cause an action to be judged as unusual, in which case the reminder acts as an implicature. Thus, a related virtue of implicit communication is that it allows the observer to maintain the pretense of having already known a fact that they forgot.

7. PRACTICAL IMPLEMENTATION

Inference, both generation and understanding, is implemented as a search over actions and facts, respectively. Techniques are needed to streamline both search problems, due to the intractability of the literal brute force search implied by argmax in (1)–(2). Existing implementations of instances of implicit communication employ AI search-pruning techniques [22, 34] or restrict the action space A_f in order to narrow the set of options under consideration [10, 29]. In practical terms, the set of feasible actions A_f is typically hard-coded for a domain, raising the possibility that it mismatches with some human's expectation. Two people may similarly encounter a mismatch in expectation about A_f. Interestingly, the machinery described in this paper could be used by a robot to infer that an observed human action is intended to accomplish a (surprising) functional goal by leveraging the context, leading to extension of A_f.

Another challenge is to build M_{pub}, the common ground model among agents. A complete model is often both unnecessary (since many facts in the agents' shared knowledge are irrelevant for the joint activity at hand) and infeasible (since the task of modeling the full common ground presents a high cognitive burden). As a result, M_{pub} need only consist of the facts that are pertinent to the success of the joint activity. For example, in the social navigation of Mavrogiannis and Knepper [29], M_{pub} might contain an updated belief regarding the destinations and intentions of observed agents. M_{pub} is therefore instantiated as the mutual understanding that the agents involved intend to participate in the joint activity along with shared knowledge about the kinds of actions that agents will likely take to contribute to the activity [4].

For humans, M_{pub} does not necessarily include all task-relevant facts at the start of the activity. It is frequently less costly to repair a misunderstanding that results from not sharing a piece of information than to expend the effort required to ground that piece of information through the principle of least collaborative effort [7, 30]. M_{pub} is then updated interactively throughout the course of the joint activity, either when new information about the intents of the agents becomes publicly available or when the agents issue a repair that helps align their own mental models of the situation (and in doing so adds to the common ground) [26]. Machine-interpretable ontologies using tools like RDF and OWL address the general problem of managing and searching M_{pub}, as exemplified by the KnowRob project of Tenorth and Beetz [33].

Finally, the distribution $P(a|M)$ is generally best modeled through machine learning. The particular context in which one takes an action affects the probabilities of observing various possible actions, often in complex ways. For example, Knepper et al. [22] employ Tellex's generalized grounding graph (G^3) [32]. Based on a conditional random field, G^3 employs a set of *correspondence variables* to valuate the correspondence probability of a given language phrase and grounding concept. These learned relationships capture concepts including objects, actions, and spatial relations.

8. DISCUSSION

Conversational implicature and legibility, though originating in different domains, are connected by techniques of encoding and decoding meaning using teleological inference [8]. 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. Several authors [14, 34] show promising early results in modeling a simple form of implicature and performing inference by model inversion.

8.1 A Call to Action

In the coming years, modeling of implied meaning, including through implicature and legible motion, will become an increasing focus within robotics – not least because humans already use these forms of implicit communication on robots today. Humans are also already interpreting robots' actions through the lens of implicit communication. Since few robots are cognizant of the implicit meaning of their actions, today's robots send random signals to humans. By and large, humans are unable to interpret robot actions in the purely functional manner that they are intended. Thus, the robotics research community must find techniques to efficiently generate and understand implicit communication.

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; today, 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. These processes are not yet understood in humans, but they will need to be deployed on robots in order to promote responsive behavior and avoid major misunderstandings.

Acknowledgments

This material is based upon research supported by the Office of Naval Research under Award Number N00014-16-1-2080 and by the National Science Foundation under Grant No. 1526035. We are grateful for this support.

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] J. M. Barbalet. "Social emotions: confidence, trust and loyalty". In: *International Journal of Sociology and Social Policy* 16.9/10 (1996), pp. 75–96.

[3] J. S. Birman. *Braids Links And Mapping Class Groups*. Princeton University Press, 1975.

[4] M. E. Bratman. "Shared cooperative activity". In: *The philosophical review* 101.2 (1992), pp. 327–341.

[5] E. Cha, A. D. Dragan, and S. S. Srinivasa. "Perceived robot capability". In: *Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on*. IEEE. 2015, pp. 541–548.

[6] H. H. Clark. *Using Language*. Cambridge University Press, May 1996.

[7] H. H. Clark and D. Wilkes-Gibbs. "Referring as a collaborative process". In: *Cognition* 22.1 (1986), pp. 1–39.

[8] 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.

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

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

[11] R. Fagin, J. Y. Halpern, Y. Moses, and M. Vardi. *Reasoning about knowledge*. MIT press, 2004.

[12] E. Goffman. *Relations in Public*. Penguin, 2009.

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

[14] 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.

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

[16] B. Hayes and B. Scassellati. "Discovering task constraints through observation and active learning". In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*. IEEE. 2014, pp. 4442–4449.

[17] B. Hayes and B. Scassellati. "Effective Robot Teammate Behaviors for Supporting Sequential Manipulation Tasks". In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*. 2015.

[18] P. Hedström and C. Stern. "Rational choice and sociology". In: *The new Palgrave dictionary of economics* (2008), pp. 872–877.

[19] J. Hohwy. *The Predictive Mind*. Oxford University Press, 2013.

[20] R. A. Knepper. "On the Communicative Aspect of Human-Robot Joint Action". In: *the IEEE International Symposium on Robot and Human Interactive Communication Workshop: Toward a Framework for Joint Action, What about Common Ground?* New York, USA, Aug. 2016.

[21] R. A. Knepper and D. Rus. "Pedestrian-Inspired Sampling-Based Multi-Robot Collision Avoidance". In: *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*. Paris, France, Sept. 2012.

[22] 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.

[23] 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.

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

[25] M. K. Lee, S. Kiesler, J. Forlizzi, S. Srinivasa, and P. Rybski. "Gracefully mitigating breakdowns in robotic services". In: *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. IEEE. 2010, pp. 203–210.

[26] S. Lemaignan and P. Dillenbourg. "Mutual modelling in robotics: Inspirations for the next steps". In: *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2015, pp. 303–310.

[27] J. Mainprice and D. Berenson. "Human-robot collaborative manipulation planning using early prediction of human motion". In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*. IEEE/RSJ. 2013, pp. 299–306.

[28] C. I. Mavrogiannis and R. A. Knepper. "Decentralized Multi-Agent Navigation Planning with Braids". In: *Proceedings of the Workshop on the Algorithmic Foundations of Robotics*. San Francisco, USA, Dec. 2016.

[29] 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.

[30] M. J. Pickering and S. Garrod. "Toward a mechanistic psychology of dialogue". In: *Behavioral and brain sciences* 27.02 (2004), pp. 169–190.

[31] 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.

[32] S. A. Tellex et al. “Understanding natural language commands for robotic navigation and mobile manipulation”. In: *AAAI Conference on Artificial Intelligence*. AAAI Publications, 2011.

[33] M. Tenorth and M. Beetz. “KnowRob: A knowledge processing infrastructure for cognition-enabled robots”. In: *The International Journal of Robotics Research* 32.5 (2013), pp. 566–590.

[34] A. Vogel, C. Potts, and D. Jurafsky. “Implicatures and Nested Beliefs in Approximate Decentralized-POMDPs.” In: *ACL (2)*. Citeseer. 2013, pp. 74–80.

[35] R. Wilcox, S. Nikolaidis, and J. Shah. “Optimization of temporal dynamics for adaptive human-robot interaction in assembly manufacturing”. In: *Robotics Science and Systems VIII* (2012), pp. 441–448.

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