Toward a model for handling noise in human-robot communication

Elizabeth I. Sklar and Elizabeth Black

Department of Informatics, King's College London, UK
{elizabeth.sklar,elizabeth.black}@kcl.ac.uk

Abstract. Human-robot interaction necessarily involves some means for the parties to communicate with each other. Whereas agent-to-agent communication is typically facilitated through electronic messaging and an agreed-upon language, communicating with humans introduces sources of noise that can disrupt the interaction in multiple ways. Consider that a human has an idea in her head, which she wishes to communicate to someone else. She must first encode this idea using some kind of language, and then transmit it. Does the receipient acquire the complete message and understand the meaning that was intended to be conveyed? The work presented here extends a theoretical model of *computational argumentation* designed to support human-robot interaction by incorporating a methodology for handling noise in communication. A running example illustrates the depth and complexity of the noise issue, and serves to demonstrate the potential of the proposed methodology.

1 Introduction

Consider the following scenario. Rosie is a nurse's aide robot tasked with looking after a patient, Hubert. Rosie evaluates the following rule to determine an action:

if	Hubert needs to take his medication at 10am.	(b_1)
and	It is now 10.05am.	(b_2)
and	Hubert has not taken his medication yet.	(b_3)
then	Rosie will give Hubert his medication.	(c)

Using *computational argumentation*[1] [31], we can express this rule as:

$$\{b_1, b_2, b_3\} \rightarrow c$$

where c is called the *conclusion* for the *argument* (S, c), and $S = \{b_1, b_2, b_3\}$ is called the *support* for the conclusion. In classical argumentation, given an argument (S, c) where $S = \{b_1, \ldots, b_n\}$, if we believe that each of the elements of the support $b_1, \ldots, b_n$ is *true*, then we have reason to believe that the conclusion c is also true. But what if we are unsure about the verity of a given proposition? In

[1] In the literature, computational argumentation may also be called *logical argumentation* or *formal argumentation*.

earlier work [34], Sklar & Azhar used the notation $?b_i$ to represent the situation where we do not know whether a belief, b_i, is true or false. If Rosie has not spoken with or observed Hubert, then she would not know whether he has taken his medication or not, and she would not be able to determine whether to take action c or not[2]:

$$\{b_1, b_2, ?b_3\} \rightarrow ?c$$

In general, the presence of any $?b_i$ in Rosie's set of beliefs is an indication that information should be gathered about b_i, particularly if a decision hinges on knowing the verity of b_i. In our example, Rosie believes that Hubert knows b_3, whether he took his medication or not. Using the human-robot argumentation framework of [34], we represent the following:

$R.\Sigma$	= Rosie's set of beliefs
$R.\Sigma \ni ?b_3$	= Rosie's lack of knowledge about b_3
$R.\Gamma(H)$	= Rosie's beliefs about Hubert's beliefs
$R.\Gamma(H) \ni (b_3 \vee \neg b_3)$	= Rosie's belief that Hubert has knowledge about b_3

From these belief stores, we deduce that Rosie should initiate an *information-seeking dialogue* [38] to obtain knowledge of b_3 from Hubert (Table 2, [34]). Following the protocol for an information-seeking dialogue (Figure 3b, [34]), Rosie will execute the first *move* in this *argumentation-based dialogue game* [20]:

$$R.\mathsf{question}(b_3)$$

to which Hubert, assuming he is following the rules of the dialogue game, will respond with one of:

$$H.\mathsf{assert}(b_3) \quad \text{or} \quad H.\mathsf{assert}(\neg b_3) \quad \text{or} \quad H.\mathsf{assert}(?b_3)$$

And the rest of the dialogue game plays out, as detailed in [34].

This example illustrates a scenario where a human and robot work cooperatively, where the robot can be proactive and can guide the human's actions [15]. Such robots must be able to justify their decisions and explain why particular actions are appropriate. In turn, humans must be able to challenge the robots' decisions and justifications; and they must be able provide input into the robot's decision-making process. It is critical that robots are able to take the *initiative* [6, 13], so that they can introduce new ideas or suggest new tasks or actions, without being prompted by a human. *Computational argumentation* is a well-studied theoretical methodology that can act as an effective, structured mechanism to allow humans and robots to offer input into one another's reasoning [24] because it formally characterises both human [22] and logic-based reasoning [8].

However, today's models of computational argumentation—like the one above which was developed to support human-robot interaction ([34]), as well as others which were developed to support agent-agent and human-agent interaction—make the assumption that the *communication* that occurs between the human

[2] We define *truth* for the ? operator as: $\mathsf{T} \wedge ? = ?$, $\mathsf{F} \wedge ? = ?$, $\neg ? = (\mathsf{T} \vee \mathsf{F})$

and the robot is perfect: if R transmits b_i, then H receives and interprets b_i, with the meaning of b_i just as R intended to convey. But as we well know, in an implemented system deployed in the real world, communication is far from perfect. There may be errors in the way a message is *encoded* (translated from an abstract thought or idea into natural language), *transmitted* (sent and received along some communication channel) and *decoded* (translated from natural language and interpreted). Such errors will likely distort the meaning of the message, where perhaps the receiver misinterprets and thus understands something very different from what the sender intended.

Consider simulated robot environments. Here, a robot has perfect knowledge of its position (location) within its environment (typically represented as some kind of "grid" world). When the robot moves, it knows exactly where it started from and where it ends up. In contrast, in physical robot environments, the job of *localisation*[3] is non-trivial due to noise in the robots' vision or other sensor systems used to estimate a robot's position. There may also be noise in the actuators, so that a *motion model* is required to estimate how far a robot moves as the result of an action. These kinds of errors due to noise compound over time. A number of probabilistic approaches have been explored for robot localisation, incorporating error models that attempt to characterise the sources of noise and improve the estimation of a robot's position. Here, we attempt to do the same thing for human-robot communication, by characterising the possible sources of error and devising a methodology for incorporating that error, probabilistically, in the communication system. We liken simulated robot environments to agent-based environments—which is where argumentation has largely developed in recent years. In agent-based environments that employ argumentation, the communication of beliefs is assumed to be perfect: if sending agent Ag_i utters b, then receiving agent Ag_j hears and understands exactly b. But in physical robot environments and in environments where robots (and agents) interact with people, the world is a noisy place.

Returning to our example, we can ask a range of questions, such as: Did Hubert hear Rosie's question correctly? Did Hubert understand Rosie's question correctly? Did Hubert formulate his answer correctly? Did Rosie hear Hubert's answer correctly? Did Rosie understand Hubert's question correctly? This paper considers these issues surrounding *noise in communication* between a human and a robot. The model presented is theoretical and aims to extend the framework that was detailed in [34]—and implemented and evaluated in [1]—here incorporating a methodology for handling noise in human-robot communication. This extension has yet to be implemented, but the theory proposed here serves as a guideline for ongoing work (see Section 5). The remainder of this paper is organised as follows. Section 2 provides a precise definition of how our model considers *noise* in human-robot communication. Section 3 reviews background

[3] Localisation in robotics refers to the task of determining where the robot is in its environment; for example, this might be represented as an (x, y) coordinate in a 2-dimensional discretised world or as $(x \pm \epsilon_x, y \pm \epsilon_y)$, where (ϵ_x, ϵ_y) represents some level of uncertainty in the robot's position.

on *computational argumentation* and *argumentation-based dialogue*, respectively, within the context of human-robot interaction. Section 4 describes our methodology for handling noise in human-robot communication. Finally, we close with a summary and outline ongoing and future work.

2 Communication and Noise

Typically, interaction between humans and robots is predicated on some form of intentional *communication*—a means for participants to send *messages* to each other. There are many different modes of communication, and we describe a taxonomy of properties that distinguishes between modes, summarised in Table 1. Messages in human-robot interaction may be **transmitted** in *open* ways, such as broadcast, where the **recipients** may be any other parties within range. Messages can also be transmitted in *restricted* ways, such as directed to specific agents, using peer-to-peer mechanisms, or particular groups of agents, using mechanisms like subscription-based forums. The **medium** for communication may be written (e.g., a document or an image), oral (e.g., speech) and/or gestural (e.g., hand signals). The means of **transmission** may be ambient (e.g., speech, gestures), physical (e.g., paper) or electronic (e.g., MP3, TXT or PDF file). Participants in the dialogue may be **located** in the same physical space (co-located) or not (remote). Thus, a particular communication mode can be described as a tuple:

$$\langle recipients, medium, transmission, locations \rangle$$

for example, sending email to a colleague in another office is:

$$\langle \textsf{peer-to-peer}, \textsf{written}, \textsf{electronic}, \textsf{remote} \rangle$$

recipients	*medium*	*transmission*	*locations*
peer-to-peer ($n = 1$)	oral (speech)	ambient	co-located (shared)
forum ($n > 1$)	gestural (signal)	physical	remote (not shared)
broadcast ($n > 1$)	written (document)	electronic	

Table 1. Properties of communication modes. Categories are listed in columns, and possible values for each property category are listed in rows within each column.

The long-term aim for human-robot communication is to employ *natural language*, though most applications today use scripts and/or some type of constrained vocabulary and message format (e.g., *controlled* or *simplified* English [3, 9, 17, 18, 26, 30]). The issues we identify in this paper focus on abstracted forms of message content. We do not model specific aspects of natural language or non-verbal forms of communication, such as gestures, facial expressions and body language (though others in the HRI and Virtual Agents communities do look at

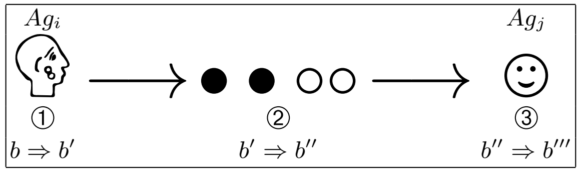

Fig. 1. Sources of noise in communication. Agent Ag_i encodes a message (①); agent Ag_i transmits the message to agent Ag_j (②); agent decodes the message (③).

these aspects). Instead, we consider communication as the transmission of *meaning* from one agent to another (human or robot). Thus, in our abstract model, the notion of *mis-communication*, errors or noise in communication, could be caused by poor choice of vocabulary, misleading hand signals, static-laden audio channels, etc. Our model generalises to cover any of the above.

2.1 Noise

Our focus is on incorporating in our model multiple forms of *noise* that may occur in human-robot communication. Figure 1 illustrates possible sources of noise in a 3-step communication process in which agent Ag_i sends a message to agent Ag_j about belief b: (1) $Ag_i.\mathsf{encode}(b)$; (2) $Ag_i.\mathsf{transmit}(Ag_j, b)$; and (3) $Ag_j.\mathsf{decode}(b)$. The first step refers to the process by which the sending agent, the "speaker", translates her belief into a language for message transmission: there may be errors in the way this encoding occurs (e.g., poor choice of vocabulary). The second step refers to the transmission of the message across a communication medium. The third step refers to the recipient decoding meaning from the message that was received.

Noise in message transmission (step ②) depends on the communication mode. For example, when communicating via physical medium in the same location, i.e., $\langle$ forum, oral, physical, co-located $\rangle$, noise in the message transmission is affected by distance between participants, ambient noise in the room and whether individuals suffer from hearing loss.

Noise in message interpretation (steps ① and ③) is always possible and may be caused by the speaker's inaccurate model of language or of the intended receiver's language abilities, which can be manifest in differences in the ontologies of the speaker and a receiver. In the case that the speaker or receiver is a robot, noise may be caused by the robot's natural language processing or speech generation capabilities. Noise in message interpretation is less likely when the message format employed is more constrained.

2.2 Modelling noise probabilistically

Given the three possible sources of noise outlined above, we consider the possibility of incremental changes to a belief, b, as it makes its way from the sender's mind to the receiver (see Figure 1):

$$\begin{aligned}\mathsf{encode}(b) &\Rightarrow b' = b + \epsilon\\ \mathsf{transmit}(b') &\Rightarrow b'' = b' + \sigma\\ \mathsf{decode}(b'') &\Rightarrow b''' = b'' + \delta\end{aligned}$$

where ϵ is the encoding error; σ is the transmission error; and δ is the decoding error. It is useful to represent each of these possible sources of error separately because, as described above, there are different reasons for error in each case.

We model the *probability* that error occurs in each case (i.e., $Pr(\epsilon)$ is the probability that there is an encoding error, etc.). These are *independent probabilities* because the presence of one type of error does not *cause* another type of error[4] (i.e., an encoding error does not *cause* a transmission error, etc.). If there is an error in encoding (b'), it is likely that the decoded message (b''') will not match the original (b); but the cause of that error will originate with the encoding process, not the decoding process. This model allows that there could also be additional decoding errors; but the occurrence of any error specific to the decoding process is not conditional on a previous error. This reasoning applies to all three types of error.

For example, suppose that there is a 90% probability of error in transmission, but only 10% probability of error in each of the remaining three steps. Thus we would have: $Pr(\epsilon) = Pr(\delta) = 0.10$ and $Pr(\sigma) = 0.9$. To determine the probability that b''' differs from b, we would compute the probability of no error, as follows:

$$Pr(b = b''') = (1 - Pr(\epsilon)) * (1 - Pr(\sigma)) * (1 - Pr(\delta)) \tag{1}$$

Working this out for our example, we have the probability that there is no error:

$$Pr(b = b''') = (1 - 0.1) * (1 - 0.9) * (1 - 0.1) = 0.081$$

and the probability that there is error:

$$Pr(b \neq b''') = 1 - Pr(b = b''') = 0.919$$

Of course, this is high because the probability of transmission error is high (90%).

Now that we have a model for representing the different types of error and how that error propagates when a message is communicated, we move on to describe how this model is used to extend the argumentation-based framework for human-robot interaction. We employ computational argumentation and argumentation-based dialogue because these are well-studied techniques that provide a solid foundation for reasoning and for shared decision making. The

[4] Of course, the different types of errors might be *correlated*; but *causation* and *correlation* are two different things, and we use independent probabilities here because there is no causation relationship between types of communication errors.

advantage of computational argumentation over other methods of negotiation is that argumentation defines a verifiable structure for associating evidence with conclusions. Thus argumentation is a natural methodology to select for interaction with humans, because humans will often ask "why" a decision is taken; and a robot that makes decisions using argumentation will easily be able to provide the evidence that supports its decisions.

3 Argumentation-based Dialogue

Much of the existing work on *argumentation-based dialogue* is classified according to Walton and Krabbe's influential two-party dialogue typology [38], which categorises dialogue types according to an *initial state* and the *dialogue goals* of the participants. We detail three types of dialogue here[5]: *information-seeking* [38], where one agent asks a question that it believes another agent knows the answer to; *persuasion* [29], where one agent tries to alter the beliefs of another agent; and *inquiry* [19], where two agents collaboratively seek the answer to a question that neither knows the answer to. Following [34], we apply these models to human-robot dialogue. In this section, we begin by explaining our terminology and reviewing some key elements of *computational argumentation* and then delve into aspects of noise and its potential impact on *argumentation-based dialogue*.

3.1 Computational Argumentation: Terminology

Computational argumentation is a well-studied form of reasoning that explicitly identifies both the justifications for any *claim* that is made by a particular argument, called its *support*, and any *conflicts* that exist between arguments [31, 2]. One can represent a set of arguments and the conflicts, or *attacks*, between them as a directed graph (referred to as an *argumentation framework*); such a graph can then be evaluated according to one of a range of *argumentation semantics*, in order to determine which arguments it is coherent to accept [8]. These semantics are based on the intuitive principles that it is not rational to accept any two arguments that are in conflict with each other, and that an argument which is attacked can only be accepted if all of its attacking arguments are themselves attacked by an accepted argument. Argumentation thus provides an intuitive mechanism for dealing with inconsistent, uncertain and incomplete knowledge, and, through structured *argument dialogues*, supports intelligent agents in reaching agreements, making decisions and resolving conflicts of opinion [24].

A robot maintains a (possibly inconsistent) set of *beliefs*, $R.\Sigma$, which is a set of *wff* of some logical language $\mathcal{L}$. We assume that $\mathcal{L}$ is sufficiently expressive to represent intentions to perform actions, as well as beliefs about the state of the world. The robot also maintains a model of what it believes are the beliefs of others with whom it interacts. We use the notation of [33] to express this notion: $R.\Gamma(H)$ represents the robot's beliefs about the human's beliefs. Note

[5] Our model could also be applied to other dialogue types from the literature.

	pre-conditions	*post-conditions*
information-seeking	$b \notin R.\Sigma \wedge b \in R.\Gamma(H)$	$b \in R.\Sigma \wedge b \in R.\Gamma(H)$
persuasion	$b \in R.\Sigma \wedge b \notin R.\Gamma(H)$	$b \in R.\Sigma \wedge b \in R.\Gamma(H)$
inquiry	$b \notin R.\Sigma \wedge b \notin R.\Gamma(H)$	$b \in R.\Sigma \wedge b \in R.\Gamma(H)$

Table 2. Dialogue protocols

that we have built our model from the robot's perspective (the entity that we can control) and do not attempt to model the human's beliefs explicitly (i.e., $H.\Sigma$ or $H.\Gamma(R)$)[6]. Following [28], the robot's *commitment store (CS)* contains locutions uttered in dialogues with the human (coming from either participant), represented as $R.CS(H) \subseteq \mathcal{L}$. As in [34], we view an agent's commitment store, $Ag.CS$, as containing the agent's *public knowledge*, since it contains information that is shared with other agents through the dialogue (like a chat log).

We assume that agents use an argumentation formalism and semantics to determine *justified claims* from their beliefs:

$$\mathsf{justified}(\Sigma) = \{b \in \mathcal{L} \mid \mathsf{acceptable}(b, \Sigma)\}$$

meaning that b is acceptable given Σ, according to the chosen argumentation formalism and semantics. We make no prescription as to which formalism and semantics are used, asserting only that the set of justified claims is consistent. There are many formalisms and associated *acceptability semantics* that may be used. For example, $\Sigma \subseteq \mathcal{L}$ may represent an ASPIC+ knowledge base [23] and we may specify that b is acceptable if and only if b is the claim of an admissible argument from the Dung-style argument framework [8] constructed from Σ and a particular ASPIC+ argumentation system.

3.2 Dialogue: Protocols and Axiomatic Semantics

Argumentation-based dialogues are typically specified in terms of *protocols*, establishing *pre-conditions* (*initial state*) and *post-conditions* (*dialogue goals*), and *axiomatic semantics* [21]: the *moves* that can be made and the *rules* that determine when moves can be made and what effect making a move has on the dialogical commitments of the participants. Table 2 shows the protocols for the three dialogue types described in [35, 36][7]. Once the pre-conditions have been met for a particular type of dialogue, then the participant initiating the dialogue can open with a move as prescribed by the axiomatic semantics defined for that dialogue. For example, a persuasion dialogue opens with an assert move uttered by the participant who initiates the dialogue. The hearer can than either accept the belief that was asserted, $\mathsf{challenge}$ the belief, or assert the opposite belief (i.e., $\mathsf{assert}(\neg b)$ in response to $\mathsf{assert}(b)$). Table 3 illustrates two sample moves,

[6] This is because we have no way of knowing exactly what the human is really thinking.

[7] Note that our formal notation differs from [35, 36]: our $R.\Sigma$ is equivalent to their Δ_R and our $R.\Gamma(H)$ is equivalent to their $\Gamma_R(H)$.

move	*pre-conditions*	*post-conditions*
$R.\mathsf{assert}(b)$	1. $b \in R.\Sigma$ 2. $(S, b) \in \mathcal{A}(R.\Sigma)$ 3. $b \notin R.\Gamma(H)$	1. $R.CS \leftarrow \mathsf{assert}(b)$
$H.\mathsf{accept}(b)$ *(terminates dialogue)*	1. $b \notin R.\Gamma(H)$ 2. $b \in R.CS$ 3. $b \in R.\Sigma$ 4. $(S, b) \in \mathcal{A}(R.\Sigma)$	1. $H.CS \leftarrow \mathsf{accept}(b)$ 2. $R.\Gamma(H) \leftarrow \{b\}$ 3. $\mathcal{A}(R.\Gamma(H)) \leftarrow \{(S, b)\}$

Table 3. Example dialogue moves. The notation $(S, b) \in \mathcal{A}(R.\Sigma)$ means that the agent R's knowledge base contains an *argument* with support S and conclusion b.

employing the axiomatic semantics outlined in adapted as if executed by Rosie (assert) and Hubert (accept). For full details on the protocols and axiomatic semantics for the three dialogue types listed in Table 2, the reader is referred to [34]; and for additional dialogue types to [35, 36].

4 An Argumentation-based Communication Model that incorporates Noise

Our aim is to represent the impact of *noise* from the communication on the interactions between the human and the robot. For our model, we have identified three key aspects of the process of using argumentation-based dialogue for human-robot interaction where noise will come into play: (1) *belief maintenance*—where the robot maintains its sets of beliefs during and after a dialogue; (2) *dialogue choices*—where the robot decides whether to initiate a dialogue and which moves to select during a dialogue; and (3) *argument acceptability*—where the robot determines the justified claims for a set of arguments. We describe each in turn, below.

4.1 Belief maintenance

As described in the previous section, we represent the robot's beliefs in multiple partitions: $R.\Sigma$ contains the robot's private beliefs about itself, others and the world; $R.\Gamma(H)$ contains the robot's beliefs about what the human believes; $R.CS$ contains a record of the utterances made by the robot during their dialogue; and $H.CS$ contains a record of the human's utterances. These belief structures are maintained using different sets of rules for each partition, as described below.

R.CS and **H.CS**. The commitment stores (or chat log), $R.CS$ and $H.CS$, contain a record of the locutions that have been issued by both participants in the dialogue. Every time an utterance is made, the commitment store is incremented to contain this new statement; thus the commitment store is updated during a dialogue. The commitment store encapsulates the messages that are transmitted, after they are encoded. For example, if the robot believes b, then $b \in R.\Sigma$.

When R transmits a message about b, incorporating the encoding error, then $R.CS \leftarrow b'(= b + \epsilon)$.

R.Γ(H). Because we do not model the human's beliefs explicitly (i.e., $H.\Sigma$), we use $R.\Gamma(H)$ as a proxy for the human's beliefs—i.e., what the robot believes the human believes. We carry the same logic (above) to the human's commitment store. If the robot hears x from the human, then we have $H.CS \leftarrow x$. Our model of noise assumes that x is an encoded version of the belief stored in $H.\Sigma$, without any encoding noise, i.e., $x - \epsilon$. Using notation consistent with above, this means that we have: $b' \leftarrow H.CS$ because $b \in R.\Gamma(H)$, where $b = b' - \epsilon$.

R.Σ. The robot's private beliefs are only updated after a dialogue has ended. Depending on the argumentation formalism and semantics in use, different rules apply for maintaining this belief partition. The main assertion here is that the robot maintains a set of consistent justified claims. This means that the robot cannot believe b and $\neg b$ at the same time, nor can it believe b and $?b$ at the same time, nor $\neg b$ and $?b$. This is where the noise factor adds a complication. For example, what is the relationship between b and b'''? The update methodology first estimates $\Delta b = b - b'''$. We set a threshold, $\mathcal{T}$, such that if $\Delta b < \mathcal{T}$, then b is close enough to b''' that they are considered the same. Thus:

$$\begin{aligned} \text{if } (b \in R.\Sigma) \wedge (\Delta b \geq \mathcal{T}) : \ & R.\Sigma \leftarrow \{b'''\} \\ \text{else} : \ & R.\Sigma \text{ does not change} \end{aligned} \tag{2}$$

Although, by the definition of our noise model, we do not know what b is when we have received and decoded a transmission as b''', we can estimate b using the probabilistic model described in Section 2.2 (obviously assuming we know the probabilities of each of type of error, ϵ, σ and δ[8]).

4.2 Dialogue choices

The decisions about which dialogue to initiate and which moves to make during a dialogue are based on the evaluation of pre-conditions, outlined in Section 3.2. As shown in Table 2, the pre-conditions are based on evaluating the membership of b in $R.\Sigma$ and $R.\Gamma(H)$. Here we adapt the rules expressed in Equation 2. We compute the *nearest neighbours* of b in $R.\Sigma$ and $R.\Gamma(H)$: $\mathsf{nn}(b, R.\Sigma)$ and $\mathsf{nn}(b, R.\Gamma(H))$, respectively. If the distance between b and each nearest neighbour is less than the threshold $\mathcal{T}$, then we consider that b belongs to each set. Assuming that beliefs are either quantitative or qualitative, we can borrow any of a number of sophisticated techniques from data mining and text mining to compute the distance between beliefs, using scoring mechanisms that attempt to quantify similarity between qualitative values. Thus we modify the rules expressed in Table 2 accordingly. The results are shown in Table 4.

For deciding which moves to select during a dialogue, the same concept applies. We leave the extrapolation of the specifics, combining this thresholding

[8] These values can be learned, but describing the steps for doing so is beyond the scope of this paper.

$$\begin{array}{l}
\text{if } (\mathsf{nn}(b, R.\Sigma) < \mathcal{T}) \wedge (\mathsf{nn}(b, R.\Gamma(H)) < \mathcal{T}) : \text{agreement (no dialogue)} \\
\text{if } (\mathsf{nn}(b, R.\Sigma) \geq \mathcal{T}) \wedge (\mathsf{nn}(b, R.\Gamma(H)) < \mathcal{T}) : \text{information seeking} \\
\text{if } (\mathsf{nn}(b, R.\Sigma) < \mathcal{T}) \wedge (\mathsf{nn}(b, R.\Gamma(H)) \geq \mathcal{T}) : \text{persuasion} \\
\text{if } (\mathsf{nn}(b, R.\Sigma) \geq \mathcal{T}) \wedge (\mathsf{nn}(b, R.\Gamma(H)) \geq \mathcal{T}) : \text{inquiry}
\end{array}$$

Table 4. Dialogue selection, using nearest neighbours

concept with axiomatic semantics from prior work (as cited in Section 3.2, to the details required for implementation.

4.3 Argument acceptability

The third aspect of our model is the consideration of the *acceptability* of arguments within the context of noise. There is a growing literature on the notion of *probabilistic argumentation*, and here we consider the equation-based approach of [10]. This method adapts the *labelling* approach of [5], whereby arguments are assigned labels as follows: **out** is for arguments that have at least one *attacker*; **in** is for arguments where all attackers are **out**; and **und**(ecided) is for arguments that have no attackers that are **in** and at least one attacker that is **und**. As outlined in [10], the probabilistic interpretation of the labels equates to: 0 for arguments that are labelled **out**, 1 for arguments that are labelled **in**, and a value between 0 and 1 for arguments that are labelled **und**.

Thus, using the probabilistic annotation of beliefs, as outlined in Section 2.2, we can evaluate a label for each argument. Then we employ the above method to justify claims and determine the acceptability of an argument, given the framework of multiple arguments.

5 Summary

A wide range of approaches have addressed the significant challenges in human-robot communication [11, 12]. Prior work where robots were deployed as receptionists [16], museum guides [37] and classroom tutors [7] largely relied on scripted dialogues to enable human-robot communication. However, a scripted dialogue model is not flexible or scalable for most situations where humans and robots work together to achieve a common goal. Thus the suggestion to employ *computational argumentation-based dialogue* as a means to support more flexible interaction is warranted.

Here, we have introduced a novel model for incorporating noise in argumentation-based communication between humans and robots. This model builds on our prior work devising argumentation-based dialogue to support human-robot interaction and investigation of different types of dialogue to support collaborative decision making. Future work will consider this model in the context of environments with multiple robots and humans [4, 27, 25] and will extend our current model of the “other” to incorporate mental state [14] and personality type [27, 32]. Our next steps with this work involve implementation and experimental evaluation of our model with a deployed robot receptionist, situated in a noisy lobby.

References

1. Azhar, M.Q.: Toward an argumentation-based dialogue framework for human-robot collaboration. Ph.D. thesis, City University of New York (2015)
2. Bench-Capon, T., Dunne, P.E.: Argumentation in artificial intelligence. Artificial Intelligence 171(10-15), 619–641 (2007)
3. Bernstein, A., Kaufmann, E., Göhring, A., Kiefer, C.: Querying Ontologies: A Controlled English Interface for End-Users. In: The Semantic Web (ISWC). vol. LNCS 3729, pp. 112–126. Springer (2005)
4. Black, E., Sklar, E.I.: Computational argumentation to support multi-party human-robot interaction: Challenges and advantages. In: Groups in HRI Workshop at the IEEE ROMAN Conference (2016)
5. Caminada, M.W.A.: On the issue of reinstatement in argumentation. In: Proceedings of the European Conference on Logic in Artificial Intelligence. pp. 111–123 (2006)
6. Carbonell, J.R.: Mixed-Initiative Man-Computer Instructional Dialogues. Tech. Rep. 11399, Bolt Beranek and Newman, Inc (BBN) (1971)
7. Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., Nabais, F., Bull, S.: Towards Empathic Virtual and Robotic Tutors. In: Artificial Intelligence in Education. vol. LNCS 7926, pp. 733–736 (2013)
8. Dung, P.M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. Artificial Intellegence 77, 321–357 (1995)
9. Fuchs, N.E., Schwitter, R.: Attempto Controlled English (ACE). In: Proceedings of the Workshop on Controlled Language Applications (CLAW) (1996)
10. Gabbay, D.M., Rodrigues, O.: Probabilistic Argumentation: An Equational Approach. Log. Univers. 9, 345–382 (2015)
11. Goodfellow, I.J., Koenig, N., Muja, M., Pantofaru, C., Sorokin, A., Takayama, L.: Help me help you: Interfaces for personal robots. In: Proceedings of Human-Robot Interaction conference (HRI). pp. 187–188. IEEE (2010)
12. Hoffman, G., Breazeal, C.: Collaboration in human-robot teams. In: Proceedings of the AIAA 1st Intelligent Systems Technical Conference (2004)
13. Horvitz, E.: Principles of Mixed-Initiative User Interfaces. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. pp. 159–166. ACM (1999)
14. Imamura, Y., Terada, K., Takahashi, H.: Effects of Behavioral Complexity on Intention Attribution to Robots. In: Proceedings of the Third International Conference on Human-Agent Interaction (HAI). pp. 65–72. ACM (2015)
15. Jennings, N.R., Moreau, L., Nicholson, D., Ramchurn, S., Roberts, S., Rodden, T., Rogers, A.: Human-agent collectives. Communications of the ACM 57(12), 80–88 (2014)

16. Kirby, R., Broz, F., Forlizzi, J., Michalowski, M.P., Mundell, A., Rosenthal, S., Sellner, B.P., Simmons, R., Snipes, K., Schultz, A., Wang, J.: Designing Robots for Long-Term Social Interaction. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1338–1343. IEEE (2005)
17. Kirk, N.H., Nyga, D., Beetz, M.: Controlled natural languages for language generation in artificial cognition. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 6667–6672 (2014)
18. Lincoln, N.K., Veres, S.M.: Natural Language Programming of Complex Robotic BDI Agents. Intelligent Robot Systems 71, 211–230 (2013)
19. McBurney, P., Parsons, S.: Representing epistemic uncertainty by means of dialectical argumentation. Annals of Mathematics and Artificial Intelligence 32(1–4), 125–169 (2001)
20. McBurney, P., Parsons, S.: Dialogue games in multi-agent systems. Informal Logic 22(3), 257–274 (2002)
21. McBurney, P., Parsons, S.: Dialogue games for agent argumentation. In: Rahwan, I., Simari, G. (eds.) Argumentation in Artificial Intelligence, pp. 261–280. Springer (2009)
22. Mercier, H., Sperber, D.: Why do humans reason? Arguments for an argumentative theory. Behavioral and Brain Sciences 34(02), 57–74 (2011)
23. Modgil, S., Prakken, H.: A general account of argumentation with preferences. Artificial Intelligence 195(0), 361 – 397 (2013)
24. Modgil, S., Toni, F., Bex, F., Bratko, I., Chesñevar, C.I., Dvořák, W., Falappa, M.A., Fan, X., Gaggl, S.A., García, A.J., et al.: The added value of argumentation. In: Ossowski, S. (ed.) Agreement Technologies, pp. 357–403. Springer Netherlands (2013)
25. Nakanishi, H., Nakazawa, S., Ishida, T., Takanashi, K., Isbister, K.: Can Software Agents Influence Human Relations? Balance Theory in Agent-mediated Communities. In: Proceedings of the Second International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS) (2003)
26. O'Brien, S.: Controlling controlled english: An analysis of several controlled language rule sets. In: Proceedings of the Joint Conference of the International Workshop of the European Association for Machine Translation and the Controlled Language Applications Workshop (EAMT-CLAW). pp. 105–114 (2003)
27. Otogi, S., Huang, H.H., Hotta, R., Kawagoe, K.: Analysis of personality traits for intervention scene detection in multi-user conversation. In: Proceedings of the Second International Conference on Human-Agent Interaction (HAI). pp. 237–240. ACM (2014)
28. Parsons, S., Wooldridge, M., Amgoud, L.: Properties and complexity of formal inter-agent dialogues. Journal of Logic and Computation 13(3) (2003)
29. Prakken, H.: Formal systems for persuasion dialoge. Knowledge Engineering Review 21(2) (2006)
30. Pulman, S.G.: Controlled language for knowledge representation. In: Proceedings of the First International Workshop on Controlled Language Applications (CLAW). pp. 233–242 (1996)
31. Rahwan, I., Simari, G.R. (eds.): Argumentation in Artificial Intelligence. Springer Verlag (2009)
32. Salvit, J., Sklar, E.I.: Modulating Agent Behavior using Human Personality Type. In: Proceedings of the Workshop on Human-Agent Interaction Design and Models (HAIDM) at Autonomous Agents and MultiAgent Systems (AAMAS). Valencia, Spain (June 2012)

33. Sklar, E.I., Parsons, S.: Towards the Application of Argumentation-based Dialogues for Education. In: Proceedings of the Third International Conference of Autonomous Agents and Multi Agent Systems (AAMAS). pp. 1420–1421 (2004)
34. Sklar, E.I., Azhar, M.Q.: Argumentation-based dialogue games for shared control in human-robot systems. Journal of Human-Robot Interaction 4(3), 120–148 (2015)
35. Sklar, E.I., Azhar, M.Q., Flyr, T., Parsons, S.: A Case for Argumentation to Enable Human-Robot Collaboration. In: Proceedings of the Workshop on Argumentation in Multiagent Systems (ArgMAS) at Autonomous Agents and MultiAgent Systems (AAMAS) (May 2013)
36. Sklar, E.I., Azhar, M.Q., Parsons, S., Flyr, T.: A Case for Argumentation to Enable Human-Robot Collaboration (Extended Abstract). In: Proceedings of Autonomous Agents and Multiagent Systems (AAMAS) (May 2013)
37. Thrun, S., Beetz, M., Bennewitz, M., Burgard, W., Cremers, A.B., Dellaert, F., Fox, D., Hähnel, D., Rosenberg, C., Roy, N., Schulte, J., Schulz, D.: Probabilistic Algorithms and the Interactive Museum Tour-Guide Robot Minerva. International Journal of Robotics Research 19(11), 972–999 (2000)
38. Walton, D.N., Krabbe, E.C.W.: Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. SUNY Press (1995)