# Learning Neuro-symbolic Dialogue Strategies for Interactive Symbol Grounding

RIMVYDAS RUBAVICIUS
*The University of Edinburgh*

ALEX LASCARIDES
*The University of Edinburgh*

SUBRAMANIAN RAMAMOORTHY
*The University of Edinburgh*

**Abstract**

Interactive task learning studies situations in which a teacher interacts with a learner to help them perform a novel task in an embodied environment. To successfully interpret the teacher's utterances, the learner has to perform *interactive symbol grounding*: it must update its prior beliefs about the mapping from symbols to referents, given their visual features, each time the teacher speaks. Interactive symbol grounding is even more challenging if the learner starts out unaware of concepts that are critical to task success. In that case, the learner must use the embodied conversation to discover and adapt to unforeseen possibilities, and so must cope with a continuously expanding hypothesis space and hence a non-stationary domain model. In this paper, we propose a neuro-symbolic model for learning dialogue strategies for achieving interactive symbol grounding. In particular, we study the effects of enriching the model with symbolic reasoning that captures the valid consequences of quantifiers (e.g., *both*, *every*). Our hypothesis is that utilizing such reasoning makes interactive task learning more data efficient. We test this empirically via a task of interactive reference resolution, in which the learner must jointly learn a grounding model and a policy for querying the teacher to enhance its accuracy in grounding. Our results show that a learner that exploits such symbolic reasoning for both decision-making and grounding is more data-efficient than learners that ignore such linguistic insights.

# 1 Introduction

Consider a general-purpose robot that assists humans in their daily activities; e.g. cleaning their dwelling, making a meal, or buying groceries. During factory development, this robot will be trained to perform sensorimotor skills such as picking, placing, pouring, and folding, and it will learn a policy of which of these actions to perform given its sensory observations (Garrett et al., 2021). Once deployed, a robot will face tasks assigned by humans as part of the human-robot interaction (Bartneck et al., 2020). For example in a shop, a customer or shop worker may request what items to place in the trolley. The robot would then make decisions on what to do, using its current interpretation of its environment, based on its sensory observations.

But what happens if the user issues an instruction that makes reference to an exotic fruit, like *rambutan*, which was entirely absent from the robot's prior experience? The robot would not know how to recognise this fruit's visual features and so determine its referents. Moreover, the robot must update the structure of the domain model with this unforeseen concept, thereby expanding the set of possible domain states, which in turn demands adapting the learner's policies. Along with the introduction of unforeseen concepts, the robot must also cope with existing domain concepts undergoing unforeseen changes over time. For example, following legislation introduced in 2021, straws are no longer made of single-use plastic, and so they have changed the way they look. An autonomous system deployed in the real world thus needs to cope with such unforeseen possibilities and the need to update the domain model via evidence from interaction.

To enable an agent to discover and adapt to unforeseen possibilities after it is deployed, the human user can take the role of a *teacher* and the robot the role of a *learner*, interacting via an embodied conversation, as illustrated in Figure 1. In this and similar embodied conversations, the teacher instructs the learner to perform a task that the learner has not performed before. The information provided by the teacher's initial task instruction is not sufficient for the learner to act—the teacher does not know that the learner's domain model does not include the concept denoted by the word "rambutan", and that this word is a neologism to the learner. The missing knowledge that is required to successfully execute the desired task (picking rambutans) is acquired in the subsequent embodied exchange. But for this exchange to have the desired effects, the learner must use the speaker's message to revise or refine its conceptualisation of the domain—in this case, to acquire a mapping from the word "rambutan" to its referents, given visual observations. Ideally, this belief update should help the agent master not only the specific task that

the user requires in the moment, but it should also deploy what it just learned about this unforeseen fruit in solving subsequent decision problems.

(M1) Teacher: Put the two rambutans into the trolley.
(M2) Learner: Before that, please show me a rambutan.
(M3) Teacher: Here. *(points to a rambutan)*
(M4) Learner: Okay. *(puts the two rambutans into a trolley)*

**Figure 1:** Example embodied conversation between learner and teacher.

The framework of *Interactive Task Learning* (ITL) (Laird et al., 2017) offers a way of using natural interactions (Clark, 1996), like the one in Figure 1, to address the task of (interactive) knowledge acquisition—e.g., learning to distinguish rambutans (Figure 2b) from other fruits (Figure 2a), based on their visual features. Such a mode of communication is essential from the social perspective to make the interaction pleasant for a human teacher (Tanevska et al., 2020), as well as to enable humans to interact with an agent without knowing its underlying hardware or software.

In this article, we present a neuro-symbolic model for using embodied conversation with the teacher to jointly learn both the domain model and a policy for solving planning problems. We focus on task-oriented conversations taking place in a shared embodied environment, consisting of natural language utterances and pointing gestures (which in Figure 1 are given in *italics*.) We aim to show that a learner that exploits the logical consequences that stem from the linguistic analysis of logical words (e.g. *both*, *every*) is more data efficient, both for learning a policy of when and how to query the teacher and for solving the task at hand.

**(a)** Fruits and vegetables the agent may be asked to manipulate.

**(b)** The exotic fruit *rambutan* that the agent was not aware before.

**Figure 2:** An illustration of an interactive task learning scenario.

There is a *prima facie* reason for thinking that such reasoning will help. For example, if a teacher points to a set of objects and says "Here is every rambutan", then the truth conditions of "every" allow the learner not only to acquire positive exemplars for learning to recognise rambutans (ie, those objects the teacher pointed to), but also negative examples (all those objects in the scene that the teacher didn't point to).

Developing a learning model that discovers and adapts to unforeseen possibilities via embodied dialogue raises several challenges that many contemporary machine learning models for knowledge acquisition do not face (see Section 2 for details). First, learning must be *incremental*: the teacher will expect the learner to update its beliefs and (behaviour) policies each time the teacher imparts new information via an utterance (and so the learner should do this). Secondly, the learning algorithms must cope with the hypothesis space of possibilities being unknown, given that unforeseen concepts and changes will be discovered via the information exchange in the dialogue (e.g., the existence of a fruit called "rambutan"). Thus updating requires the learner to adapt its prior probability distributions over possibilities to a newly expanded set of possibilities when such discoveries surface.

Previously, in Rubavicius and Lascarides (2022), we developed methods for performing *interactive symbol grounding* that met these challenges, using linguistic analyses of the teacher's assertions, and in particular the truth conditions of quantifiers, to inform belief update about the mapping from symbols to denotations, given sensory observations. This prior work demonstrated empirically the advantage of exploiting the truth conditions of quantifiers to interactive symbol grounding. But the model we studied in that earlier paper did not handle *dialogue per se*: the teacher did all the talking! This paper fixes that limitation.

In this paper, we expand our earlier line of research by endowing the learner with the capacity to *query the teacher* (e.g., M2 in Figure 1), and through that to control which observations it will acquire next— a form of active learning. But expanding the set of actions that the learner can perform to include dialogue moves yields a dilemma: should the learner query the teacher about reference (which comes at a cost because of the teacher's effort in answering), or should it use its current estimate of the domain model to execute what it thinks is a valid domain-level plan, thereby risking a large cost if the learner gets it wrong? In other words, for active learning to be effective, the learner must use evidence from its prior experience with embodied conversations to learn a policy that resolves this dilemma of when (and what) to ask vs. when (and how) to act in the domain. This learned policy must be neural-symbolic, as it has both a neural component (using visual observations

to estimate beliefs about the mapping from words to their referents, stemming from object similarity) and a symbolic component (performing logical reasoning about possible learner-teacher message exchanges). In this paper, we develop a model for learning such a policy, and we aim to show that exploiting the logical consequences of quantifiers like *both* and *every* not only makes grounding more data efficient (as our prior work showed), but it also makes learning policies that resolve the dilemma between asking vs. acting more data efficient as well.

To study the effects of neural-symbolic models that jointly learn both dialogue strategies and grounding models, we consider a task of *interactive reference resolution* in ShapeWorld (Kuhnle and Copestake, 2017): the teacher and learner share the same visual scene, and the teacher instructs the learner to point to a referent of a referential expression within that scene, such as "the one red square." The learner in this situation faces a choice. Firstly, it can choose to query the teacher (e.g. "before that, show me a red object"), and so incur the cost of this while benefiting from learning something useful, in that it improves its interpretation of the visual scene. Alternatively, it can take the risk of identifying a referent based on its current interpretation of the scene, which if wrong yields a substantial negative reward. The embodied conversation continues until the learner stops querying and identifies a referent, at which point the task ends.

The remainder of this article is structured as follows. We first review related work in both machine learning and computational linguistics, focusing on the extent to which this work meets the above challenges. We further provide a step-by-step exposition of Rubavicius and Lascarides (2022) for the interactive symbol grounding procedure, including several illustrative examples of how it affects belief update. These two sections serve a pedagogical purpose and can be skipped or read on their own.

Our first novel contribution relative to this prior work is to design a neuro-symbolic procedure that draws on the logical consequences of the teacher's assertions, as well as sensory observations, to jointly learn both a grounding model and a decision-making strategy that optimises the above exploration-exploitation dilemma in achieving accurate symbol grounding. Our model also supports neuro-symbolic learners that need to fix a deficient hypothesis space of the possible domain states, because the learner is initially unaware of domain-level concepts that are critical for task success.

The model itself is quite general, but we test it in proof-of-concept experiments (i.e., experiments that involve a relatively small domain of entities and concepts) to showcase the model's advantages over those that don't draw on logical inferences from natural language semantics. Specifically,

our second contribution in this paper is to test the procedure empirically by conducting experiments in ShapeWorld. We compare a model for grounding and decision-making that uses the logical consequences of quantifiers to a model that does not use this information. We showcase the learner's capacity to cope with a non-stationary set of possible domain states by making the learner unaware of all open-class words and the concepts they refer to at the start of its learning process.

The results show that utilising the symbolic logical reasoning borne from the semantics of quantified referring expressions leads to data efficiency, for both learning effective decision-making to achieve accurate grounding and for learning the grounder itself. These experiments bear out the intuition that such reasoning helps the learner to acquire more training exemplars to inform grounding from the teacher's utterances than learners that ignore the different semantics of different quantifiers.

# 2 Related Work

Before going into the details of our proposed model, we review related work in lifelong machine learning, symbol grounding, and reference resolution so as to distinguish and highlight the contributions of this paper.

## 2.1 Lifelong Machine Learning

Machine learning is used by autonomous systems to learn to solve planning problems, with training utilising data that's based on observing the consequences of one's own actions (Sutton and Barto, 1998) or the actions of others (Bishop, 2007). It has been an important paradigm in AI system design in recent decades (Bengio et al., 2021). The subfield known as lifelong machine learning considers systems that can learn many tasks over the system's life cycle from one or more domains. Such systems efficiently and effectively retain the knowledge they have acquired so far and use it for more efficient and effective learning of new tasks (Silver et al., 2013). A key element of this learning paradigm is that learning is *incremental*: it updates beliefs and policies and adapts its decision-making in real-time (online), as and when it gathers new evidence. This is in contrast to most contemporary learning paradigms, in which the learning signal is *batched*, and models train offline on large volumes of data (Mohri et al., 2012).

van de Ven et al. (2022) distinguishes three types of incremental learning scenarios, which applied to symbol grounding amount to the following three

cases: task-level (learn to solve distinct tasks with respect to a single probability distribution over possible environments); domain-level (learn to solve a single task in 'out-of-distribution' environments); and class-level (learn to solve distinct tasks while at the same time inferring a probabilistic representation of the environment). In all these scenarios, the central aim is the *reusability* of what an agent has learned so far when applied to novel or out-of-distribution scenarios. This has motivated research in transfer learning (Pan and Yang, 2010) (how to transfer the experience from one task to another), multitask learning (Caruana, 1997) (using the experience from multiple tasks to bootstrap the learning process) and meta-learning (Vilalta and Drissi, 2002) (learning how to learn).

Conventionally, the task structure is not explicitly used in these learning scenarios. But it is appealing to explore the compositional task structure for systematic generalization (Mendez and Eaton, 2022; Hupkes et al., 2020). The research in Interactive Task Learning (Laird et al., 2017) aims to build compositional approaches to policy learning, in which the system can reuse and combine in novel ways components that it has learned from prior tasks when facing novel tasks. According to van de Ven et al's classification, this makes ITL a class-level incremental learning problem (learning both the task and how to represent the domain).

Another distinguishing feature of ITL is that the learning scenario involves a teacher providing the learner with contextually relevant guidance or advice, either through an embodied natural language conversation or via action demonstrations. This article is particularly interested in how to utilise the teacher's embodied utterances to enhance, in a linguistically principled manner, the leaner's model of grounding—that is, the way the agent interprets its perceptual sensors, and in particular, identifies within the current visual scene the denotations of referring expressions. This paper adopts the ITL approach by exploring in detail the interaction between the agent's beliefs about grounding and the decision-making that's required to perform complex tasks, where the latter relies entirely on generic decision-making algorithms (Kochenderfer et al., 2022), but the novelty comes from utilizing the logical consequences of quantifiers during training.

Changes in perceptual capabilities can lead to two types of domain model updates: *parameter-level*, in which the underlying structure of the domain conceptualisation stays the same but its parameters are tuned to the current scenario; and *structure-level*, in which the domain conceptualisation itself changes (acquiring skills and concepts that were not previously a part of the domain at all, thereby expanding the hypothesis space of possible domain states and actions). The latter is the primary interest in this article. To update

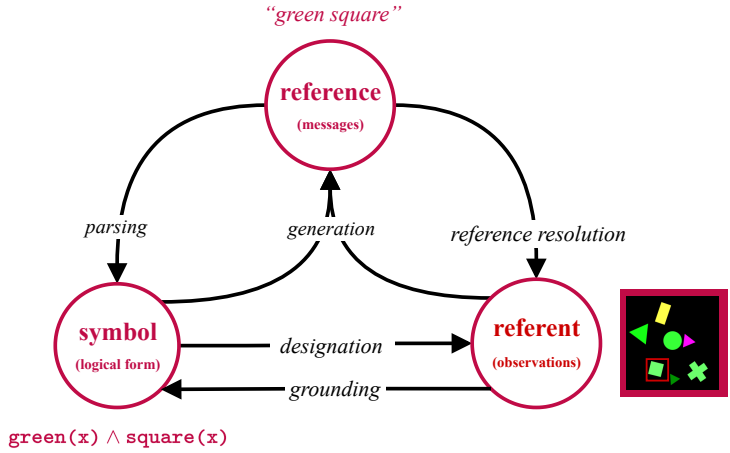



**Figure 3:** Reference triangle (Ogden et al., 1924) showing the relationship between the reference (linguistic form/messages), symbol (logical form/ground truth) and referent (entities in the environment) and commonly discussed tasks like (semantic) parsing, generation, reference resolution, grounding (inferring the domain model), and designation.

the agent's conceptualisation of the domain (and in particular its structure), we study the learning signal of the embodied conversation (Cassell, 2001) to discover and then exploit unforeseen possibilities that are critical to the task.

By design, such scenarios are not in competition with large multimodal models (Tan and Bansal, 2019; Li et al., 2023; Driess et al., 2023) that are trained offline via batch learning and are designed to support broad domain coverage and parameter-level model updates in interaction. Research in ITL aims to provide a means of performing structure-level model updates, which are needed when AI systems are deployed after training offline on 'big data' in settings where unforeseen concepts are frequently introduced and/or existing concepts change in unforeseen ways. These scenarios require models to adapt, using all possible incidental supervision signals (Roth, 2017).

## 2.2 The Symbol Grounding Problem

The symbol grounding problem is the task of learning a grounding model (grounder) that links referents (sets of entities) in the environment to abstract

concepts known as symbols, on the basis of the agent's sensory observations (and in particular the visual features) (Harnad, 1990). Figure 3 outlines the reference triangle (Ogden et al., 1924) that relates the linguistic expressions, the symbolic representations of those expressions and the environment, so as to establish the relationship between the three. Symbol grounding has been extensively studied before (Hu et al., 2016b,a; Du et al., 2021; Chandu et al., 2021), in particular in designing autonomous robotic agents (Matuszek, 2018). But for ITL scenarios, many of the contemporary grounders are not sufficient, because they don't meet all the desiderata that arise from the ITL setup. These desiderata are elicited and explained below.

Firstly, from the machine learning perspective, we are tackling *multilabel classification with an expanding hypothesis space*. It might be tempting to formalise the problem as a multiclass classification problem in which, given the observation, a single symbol is predicted (Krishna et al., 2016). But such a modelling approach assumes mutual exclusivity between symbols, which might not be the case. For example, two symbols `square` and `rectangle` are not mutually exclusive because the former is a hyponym of the latter ("is a" relation); similarly for `magenta` and `red`. Because of this, a multilabel classification approach needs to be deployed in which for each entity in the visual scene, multiple symbols can be predicted. Furthermore, the key feature of ITL is that novel symbols (neologisms) are introduced in the domain, which may refer to existing concepts or entirely new and unforeseen concepts; thus the method must handle the expanding hypothesis space of labels (or symbols).

Secondly, as mentioned in Section 1, the grounder in ITL scenarios should ideally admit *active learning* and in turn real-time knowledge acquisition for decision-making. Contemporary grounders (Ye et al., 2019; Datta et al., 2019) heavily utilize offline learning for extensive fine-tuning and batch learning, which results in data regularization for the stochastic representation of the learning grounder. Even though these design decisions lead to increased performance, such grounding models are not suitable for *interactive learning*. During the embodied conversation, the knowledge is actively acquired piecemeal from the sequentially uttered messages. This knowledge has to influence the agent's decision-making as to its next move *during* the interaction with the teacher—in effect, the learner should update its beliefs and behaviour every time the teacher says something. As a result of this, incremental and online learning methods are necessary.

Thirdly, due to data scarcity, *few-shot learning* must be performed. Even when extensively utilizing the learning signal from the embodied conversation, it is unreasonable to expect that sufficient data is acquired through in-

teraction to deploy data-intensive learning methods. Hence, few-shot learning approaches that generalize from a few learning exemplars (Wang et al., 2020b) should be used.

Finally, the method has to be able to *reason about beliefs*. Because of the strong assumptions that are necessary for efficient knowledge acquisition, it is likely that some of the agent's assumptions will bring inconsistencies in the beliefs about the underlying ground-truth domain model: e.g., the teacher's latest message is inconsistent with the learner's previously made predictions. Because of this, the method of active knowledge acquisition should be able to detect inconsistencies and be able to revise and repair beliefs accordingly (Hansson, 2022).

In this article, the symbol grounding problem with these additional requirements is referred to as an *interactive symbol grounding problem*.

Symbol grounding models are key components for more complex tasks like visual question answering (VQA) (Antol et al., 2015) or manipulating an embodied environment (Alomari et al., 2017a). For these tasks, grounding models are often composed by exploiting the principle of compositionality of natural language (Mao et al., 2019; Wang et al., 2023)—i.e., that the meaning of a phrase is a function of the meaning of its parts and how they are put together. This compositionality enables the learned models to exhibit systematic generalization (much like the lifelong machine learning scenario). They acquire a grounded natural language lexicon (Mao et al., 2021) and/or grammar rules (Alomari et al., 2017b,a) from evidence. Our work doesn't learn mappings from natural language to logical form; our experiments will assume this mapping is already known. But we exploit compositionality in another way: namely, unpacking the truth conditions of arbitrarily complex logical forms is compositional. We in essence, therefore, are exploring how the valid consequences of the meaning of a semantic representation of a natural language expression assist generalization during learning and so enhances data efficiency. In particular, exploiting formal semantic interpretations enables us to acquire *negative* examples for the concepts, as observed in (Rubavicius and Lascarides, 2022) and the example involving "every" given in Section 1. It is worth noting that such notions of compositionality are not at odds with exploiting it to aid grounded semantic parsing, but complementary.

From the symbol grounding perspective, our work is most similar to Alomari et al. (2022). They perform incremental updates of visual features to arrive at a set of concepts that are clusters of visually similar situations. But by the time their system processes language, those incrementally acquired domain concepts, on the basis of visual processing, are fixed. As a consequence, symbol grounding is restricted to mapping words to that fixed set of

domain concepts. This means that they cannot reliably distinguish lexical relationships, in particular synonyms and hyponyms, in those situations where words denote visually similar concepts: e.g., burgundy glass vs. chardonay glass vs. wine glass. In contrast, we are developing models that incrementally learn from both vision and language *simultaneously*, so that the set of domain concepts is not determined solely by visual information, but also by the way those concepts are described and referred to by the teacher. Further, because Alomari et al. (2022) do not support incremental learning from language, their models aren't suited to ITL via embodied conversation.

Finally, Dobnik et al. (2022) offers a careful study of meaning representations of embodied conversation in the context of symbol grounding. They advocate using probabilistic Type Theory with Records (TTR) (Cooper et al., 2015; Cooper, 2023) to model symbol grounding from both visual percepts (Larsson, 2013; Larsson et al., 2021) and definitions (Larsson, 2021; Noble et al., 2022; Noble and Ilinykh, 2023). TTR offers a much richer framework for representing natural language utterances than first-order logic. But in this work, we do not aim to commit to a specific framework for representing natural language meaning. Rather, our aim is to present a general neuro-symbolic reasoning framework that can later be adopted by different machine learning and linguistic researchers when they design more robust natural language interfaces, handling richer interactions as part of the embodied conversation from a linguistic and perceptual point of view. Accordingly, our model of neuro-symbolic reasoning (Manhaeve et al., 2021) utilizes components that have as broad an appeal to researchers in machine learning and linguistics as possible: specifically, the symbolic reasoning stems from first-order logic representations of natural language utterances, we deploy neural similarity-based classification, and the probabilistic calculus we use to combine these elements is the well-established weighted model counting (Chavira and Darwiche, 2008) (see Sections 3 and 4 for details of the model).

## 2.3 Reference Resolution

Reference resolution is the problem of identifying a referent of a linguistic expression. Often in experimental settings, the environment of interaction is an image or a simulated environment and the conversation includes referring expressions, occasionally accompanied by pointing gestures towards objects in the visual scene (Das et al., 2017; Haber et al., 2019; Kottur et al., 2019; Loáiciga et al., 2021). In such situations, both the visual and verbal context have to be used to resolve dialogue-specific phenomena such as coreference (Kottur et al., 2018). In this work, however, we avoid problems of

co-reference in dialogue. We treat reference resolution as the task of identifying for each referential expression in the embodied conversation the entities in the visual scene that it denotes, thereby making our task correspond to the symbol grounding problem (Section 2.2).

Signalling games (Fudenberg and Levine, 1998) are often used to model the pragmatics of dialogue, with various definitions of game equilibria used to predict what speakers choose to say and how their interlocutors interpret the speaker's (ambiguous) signals (Thompson and Kaufmann, 2010; Caelen and Xuereb, 2011; DeVault et al., 2005). The probabilistic pragmatic framework of rational speech acts (RSA) (Frank and Goodman, 2012; Goodman and Stuhlmüller, 2013; Goodman and Frank, 2016) provides a game-theoretic view of such interactions by explicitly defining the speaker and listener behaviours recursively (Andreas and Klein, 2016; Monroe et al., 2017; Zarrieß and Schlangen, 2019; White et al., 2020; Fried et al., 2021). One class of such signalling games models pedagogical teacher-learner interaction (Shafto et al., 2014; Rafferty et al., 2016), in which the initiative is given to the teacher to transfer the desired knowledge to the learner, utilizing various cues and biases (Csibra and Gergely, 2009; Jara-Ettinger et al., 2016). For instance, the teacher may designate a *representative* sample of the referents of a symbol—"blickets", say—that the learner is attempting to ground while uttering "here are some blickets." In this paper, we are considering the opposite scenario: namely, one where the initiative resides with the learner to choose a query to address to the teacher, whose response is maximally likely (given the learner's current beliefs) to help it achieve the task at hand. Most teacher-learner scenarios between humans involve initiative on both sides, and it would be interesting to explore equilibria in situations where both the teacher and the learner can take initiative on how to progress the conversation. But that is beyond the scope of this paper. We focus instead entirely on the problem of how a learner can use their experience so far to learn a decent strategy for obtaining the information they need from the teacher.

RSA and cooperative communication in general (Wang et al., 2020a; Hao et al., 2023) have shown theoretical and empirical results aligning with established observations of pragmatic implicatures and other context-sensitive aspects of dialogue interpretation (Grice, 1975a; Sperber and Wilson, 1986). Nevertheless, modelling embodied conversation between the learner and a teacher in an ITL setting as a collaborative game is insufficient. By design, such games assume perfect information about the hypothesis space: all agents share perfect knowledge about what the possible domain states are, and what the possible (dialogue) actions are. However, having this level of knowledge is not always a reasonable assumption. Asher and Lascarides (2013) argue

that it is too strong an assumption in certain strategic conversations. More generally, thanks to the information exchange that occurs in dialogue, an interlocutor may learn of domain-level possibilities that it was unaware of prior to the conversation taking place—in effect the player discovers he is playing a game with a different hypothesis space of possibilities than he thought when he started playing it. Once the unforeseen possibility is discovered, the agent must refine and extend its hypothesis space of domain states accordingly—in other words, it calls for a structure-level model update. For RSA, such scenarios are not permitted. Being built on game theory, a conversation that entertains a different (or larger) set of possible domain states and/or signals is treated as a different and unrelated game. But in ITL (and more generally in lifelong machine learning) the agent, on discovering that their hypothesis space of possible states is deficient, should *retain* at least some of what it has learned so far and use it to influence its next move in the 'larger' game, even within the current interaction.

Indeed, scenarios in which the learner has a deficient domain model are a key area of interest in ITL. Because of this, the methods developed in this article are related to but do not simply reuse the RSA framework. We share with RSA a methodology that is based on a generic information-gathering learner architecture (Russell and Norvig, 2020). There are also approaches of dialogue modelling that do not explicitly model the conversation as a game but make the speaker *listener-aware*, so as to modulate their decision-making in inference (DeVault et al., 2005; Vedantam et al., 2017) and learning (Mao et al., 2016; Yu et al., 2017) or implicitly model well-known patterns inefficient communication in human-to-human conversation like reference reuse (Takmaz et al., 2020), descriptiveness (Takmaz et al., 2022), information-density (Giulianelli et al., 2021) and discourse context (Giulianelli and Fernández, 2021). For us, we study the scenario when learning and inference are not only listener-aware but also *semantics-aware*, meaning that the learner reasons about the valid consequences of messages about the domain.

At the same time, our experiments are on synthesized embodied conversations, rather than humans conversing 'in the wild', so that we retain the control that's necessary for quantifying the difference in performance between an agent that utilizes the logical consequences of quantified expressions vs. an agent that does not. How often such insights help in real human-human conversation depends heavily on how the task at hand influences that dialogue exchange (in particular, it would need ample use of different determiners and presupposition triggers). Note that both RSA-style game-theoretic approaches and listener-aware dialogue modelling approaches discussed above have been effective to show how speakers can adapt in interaction to mis-

aligned beliefs and asymmetry in knowledge for parameter-level model update (Wang et al., 2016; Wang, 2017; Hawkins et al., 2020; Takmaz et al., 2023), but we want to support structure-level model updates (i.e., learning the set of possible concepts that define the set of possible domain-level states), not merely parameter-level updates.

Finally, it is worth noting that for reference resolution, we are using an off-the-shelf semantic parser. This contrasts with several existing works, in which both the semantic parser and dialogue strategy are jointly learned. For example, Padmakumar et al. (2017); Thomason et al. (2020) address interactive symbol grounding, focussing on understanding commands by jointly learning a semantic parser and grounder. Our focus is different: while these models learn mappings from natural language to logical forms, they do not utilise the valid consequences of those logical forms to inform the learning process. This paper aims to demonstrate that those logical consequences make learning more data efficient.

# 3 Preliminaries

Before describing in detail the neuro-symbolic model for learning a dialogue strategy to achieve reference resolution (see Section 4), we provide in this section details of the model of interactive symbol grounding—that is, how the learner uses evidence from an embodied conversation with a teacher to update its beliefs about both the set of possible domain states (which is ever expanding thanks to neologisms that occur in the conversation) and the mapping of symbols to their referents, given visual features. This interactive grounding model is essentially a recap of the one presented in Rubavicius and Lascarides (2022). The novelty in this paper is in Section 4, where the grounding algorithms from this section get combined with active learning: we specify a method that enables the learner to work out, based on its experience of conversation so far, what to do so as to acquire the evidence needed for its grounding model at minimal cost (and risk) to itself and to the teacher.

Table 9 and Table 10 in Appendix A gives a summary of the symbols and their description used in this section.

## 3.1 Reasoning about the Domain

Since we will exploit a model-theoretic semantics for interpreting natural language (Hodges, 2022) so as to learn efficiently, we start by stipulating those semantics, with a focus on referring expressions. As we mentioned earlier,

our symbolic component will be based on first-order logic. So we represent a domain model $\mathcal{M}$ as a triple $\mathcal{M} = (U, V, I)$ consisting of a set of entities $U$, the vocabulary of the domain $V$ and an interpretation function $I$ that maps predicate symbols (concepts) from a vocabulary to their denotations—sets of entities for one-place predicates (properties), sets of pairs of entities for two-place predicates (relations), and so on.

Throughout this paper, we will use the toy example domain model in Eq. 1 to illustrate various features of the learning process (our experiments in Section 5 draw on larger models than this). This consists of four entities: $u_1$, $u_2$, $u_3$,and $u_4$ and four properties: `blue`, `red`, `circle`, and `square`:

$$\begin{aligned} &\mathcal{M}: \\ U &= \{u_1, u_2, u_3, u_4\} \\ V &= \{\texttt{blue}, \texttt{red}, \texttt{circle}, \texttt{square}\} \\ I &= \{\texttt{blue} : \{u_1, u_3, u_4\}, \texttt{red} : \{u_2\}, \\ &\quad \texttt{circle} : \{u_1, u_3\}, \texttt{square} : \{u_2, u_4\}\} \end{aligned} \tag{1}$$

The domain model can be equivalently defined as the (maximal) set of true atomic formulae (atoms) $a$ that are in the Herbrand base $\mathcal{H}$: these are the formulae constructed from $n$-place predicate symbols and $n$ terms referring to entities (Chang and Lee, 1973) (without loss of generality we use the term $u_i$ to denote the entity $u_i$ in the model). We call the maximal set of atoms that are satisfied by $\mathcal{M}$ $\mathcal{H}_\mathcal{M}$; i.e., $\mathcal{H}_\mathcal{M} = \{a \in \mathcal{H} \mid \mathcal{M} \models a\}$. So the model in Eq. 1 has the Herbrand base $\mathcal{H}$ in Eq. 2 and the domain model representation $\mathcal{H}_\mathcal{M}$ in Eq. 3:

$$\begin{aligned} \mathcal{H} = \{&\texttt{blue}(u_1), \texttt{blue}(u_2), \texttt{blue}(u_3), \texttt{blue}(u_4), \\ &\texttt{red}(u_1), \texttt{red}(u_2), \texttt{red}(u_3), \texttt{red}(u_4), \\ &\texttt{circle}(u_1), \texttt{circle}(u_2), \texttt{circle}(u_3), \texttt{circle}(u_4), \\ &\texttt{square}(u_1), \texttt{square}(u_2), \texttt{square}(u_3), \texttt{square}(u_4)\} \end{aligned} \tag{2}$$

$$\begin{aligned} \mathcal{H}_\mathcal{M} = \{&\texttt{blue}(u_1), \texttt{blue}(u_3), \texttt{blue}(u_4), \texttt{red}(u_2), \\ &\texttt{circle}(u_1), \texttt{circle}(u_3), \texttt{square}(u_2), \texttt{square}(u_4)\} \end{aligned} \tag{3}$$

This alternative representation of the model will be useful when developing the learner's probabilistic model of belief.

We assume that the learner can use their visual sensors to detect all the entities in the visual scene (via object detection): in other words, we treat $U$ as observable. Furthermore, as we will shortly describe, the learner can also use their visual sensors to observe the visual features of each $u \in U$

(see Sections 3.2 and 5.2 for details). But the interpretation function $I$ is latent: not only does the learner not know which symbols in $V$ denote which entities in $U$, but perhaps more profoundly $V$ itself is latent as well, due to the learner's (initial) unawareness (it will discover new symbols in $V$ as and when the teacher utters a neologism).

Since $I$ and $V$ are latent, the learner must estimate the domain model $\hat{\mathcal{M}}$ using evidence. One type of evidence is a domain theory $\Delta$, which is a set of logical formulae $\phi$ that is satisfied by the ground truth model $\mathcal{M}$: i.e., $\mathcal{M} \models \Delta$. We will assume that the learner can accurately convert the teacher's utterances into a logical form, and moreover, the teacher knows $\mathcal{M}$ and is always sincere (so she believes what she says and everything she says is true). Under these assumptions, the learner can build $\Delta$ from the teacher's (embodied) utterances in their conversation: every time the teacher speaks (and points), asserting information that has logical form $\phi$, the learner can add $\phi$ to $\Delta$ and assume that $\mathcal{M} \models \Delta$. Hence $\hat{\mathcal{M}}$ should satisfy $\Delta$ too.

We now describe two things in sequence: first, how the learner can estimate $\hat{\mathcal{M}}$ given $\Delta$, and second the kinds of information that populate $\Delta$ in our experiments, given the task the learner has to master and the kind of utterances the teacher conveys to guide the learner (in particular, the teacher uttering a referential expression and pointing to a referent of it, as illustrated in Fig. 1).

We start with the method for estimating $\hat{\mathcal{M}}$ given $\Delta$. This task corresponds to estimating the probability that any arbitrary logical formula $\phi$ is true, given $\Delta$. The fundamental building block we use for computing such probabilistic queries is a *weighted model counting* `WMC` procedure (Chavira and Darwiche, 2008). This estimates the probability that $\phi$ is true by marginalising over all models that satisfy $\phi$, and treating each atom $a \in \mathcal{H}$ as a Bernoulli random variable and assigning a probability to each of them, referred to as weights $\mathbf{w} \in [0,1]^{\mathcal{H}}$:

$$\texttt{WMC}(\phi, \mathbf{w}) = \Pr_{\mathbf{w}}(\phi) = \sum_{\mathcal{M}\colon \mathcal{M} \models \phi} \prod_{i=1}^{|\mathcal{H}|} w_i \mathbb{1}_{a_i \in \mathcal{H}_{\mathcal{M}}} + (1 - w_i)(1 - \mathbb{1}_{a_i \in \mathcal{H}_{\mathcal{M}}}) \tag{4}$$

where $\mathbb{1}_{a_i \in \mathcal{H}_{\mathcal{M}}}$ is an indicator function which is 1 if atom $a_i$ is part of the domain model $a_i \in \mathcal{H}_{\mathcal{M}}$ and 0 otherwise. Using uniform weights $w_i = 0.5 \quad \forall w_i \in \mathbf{w}$ makes `WMC` equivalent to model counting otherwise known as a #SAT problem (Valiant, 1979). We'll see in §3.2 how reasoning about sensory (i.e., visual) observations combined with the teacher's utterances over the course of the embodied conversation affect weights and make

them deviate from 0.5. But the atoms created from neologisms are initially assigned a probability of 0.5, to reflect complete ignorance about their denotations. Moreover, since WMC considers *only* atoms created from predicates in $\phi$, the inference procedure uses only a subset of domain models that could be created from the vocabulary while implicitly marginalizing over unused predicates and entities. For example, if $\phi = \texttt{red}(u_1) \wedge \texttt{square}(u_2)$ we consider domains with $U = \{u_1, u_2\}$ and $V = \{\texttt{red}, \texttt{square}\}$ while domains with more entities and predicates are ignored. This is important in ITL because it allows us to dynamically change and expand the domain as new symbols or entities are introduced to the discourse.

This article considers two types of probabilistic queries. *Complete evidence* EVI computes the probability that formula $\phi$ is true, given $\Delta$ is true:

$$\texttt{EVI}(\phi, \Delta, \mathbf{w}) = \Pr_{\mathbf{w}}(\phi \mid \Delta) = \frac{\texttt{WMC}(\{\phi\} \cup \Delta, \mathbf{w})}{\texttt{WMC}(\Delta, \mathbf{w})} \tag{5}$$

And *maximum a-posteriori* MAP estimates the most probable domain model $\hat{\mathcal{M}}$ for which $\Delta$ holds:

$$\texttt{MAP}(\Delta, \mathbf{w}) = \hat{\mathcal{M}} = \arg\max_{\mathcal{M}} \Pr_{\mathbf{w}}\Big(\bigwedge_{a \in \mathcal{H}_{\mathcal{M}}} a \bigwedge_{a \notin \mathcal{H}_{\mathcal{M}}} \neg a \mid \Delta\Big) \tag{6}$$

Now we address the logical representations of the teacher utterances that get added to $\Delta$ over the course of the embodied conversation, and their model-theoretic semantics, whose logical consequences, in turn, affect the learner's estimates of $\hat{\mathcal{M}}$ via the above equations. In this paper, we study a scenario in which the teacher's assertions consist of a referential expression $r$ combined with pointing to a denotation of $r$ (there may be more than one denotation in the model, in which case the teacher chooses which one to point to).[1]

Linguistically, $r$ is an arbitrary complex noun phrase like *"the one red square"*. We draw on generalized quantifiers (Barwise and Cooper, 1981) to represent $r$'s logical form $\Phi(r)$. Since $r$ is a noun phrase and not a sentence, $\Phi(r)$ consists of a generalized quantifier Q (corresponding to $r$'s determiner), its restrictor $R$ (corresponding to $r$'s adjectives and nouns) but the body of the quantifier is 'missing'. This is traditionally captured using $\lambda$-calculus (i.e., $\lambda B.Qx(R(x), B(x))$ where $R$ is the restrictor and $B$ the body). But

[1] We do not consider vague quantifiers(Bradburn and Miles, 1979) like "some" or "few" which implicates soft constraints on objects not designated as a consequence of pragmatic principles of cooperative conversation(Grice, 1975a). Interpretation of such quantifiers are out of the scope, but in principle could be modelled in our probabilistic framework.

this is not quite what we want to capture for the purposes of solving reference resolution, for this $\lambda$-term denotes the set of properties satisfied by the entities that satisfy $R$, with $Q$ imposing constraints on the relationship between the denotations of $R$ and $B$. In our scenario, however, we need to identify the set of referent(s) that are denoted by the referring expression $r$, not the properties satisfied by those referents.

In view of this, we express the logical form $\Phi(r)$ of the referential expression $r$ another way and assign it a semantics that draws on that from generalized quantifier theory. Specifically, we will represent the logical form $\Phi(r)$ of the referential expression $r$ as a well-formed expression of the form $\langle \texttt{Q x}.\phi \rangle$, where `Q` is a generalized quantifier (see Table 1 first column for the quantifiers we use in our experiments in Section 5) and $\phi$ is a logical formula with only one free variable `x`, which corresponds to a representation of the natural language description. $\phi$ is constructed recursively from predicate symbols like `square`, terms consisting of variables `x` and constants `u`[2], logical connectives (disjunction $\vee$, conjunction $\wedge$, negation $\neg$).[3] For example, the referential expression *"the one red square"* is assigned the logical form $\langle \texttt{\_the\_1\_q x.red(x)} \wedge \texttt{square(x)} \rangle$.

The logical form $\Phi(r)$ of a referential expression $r$ is evaluated with respect to the domain model $\mathcal{M}$ to yield a set of sets of entities—the referent $\mathcal{R}$ of $\Phi(r)$ with respect to $\mathcal{M}$. An element of the set $\mathcal{R}$ is a *set* of entities because a denotation of $r$ may include more than one entity in $U$ (e.g., a denotation of *two squares*). $\mathcal{R}$ itself is a set of (potentially) more than one element because $\mathcal{M}$ may have more than one denotation for $r$: e.g., if there is more than one entity in $\mathcal{M}$ that's a square, then *a square* has more than one denotation in $\mathcal{M}$. For example, given the model $\mathcal{M}$ in Eq. 1, $\langle \texttt{\_every\_q x.square(x)} \rangle$ should denote $\{\{u_2, u_4\}\}$ (there is only one entity that is denoted by "every square", and that is the maximal set of entities $\{u_2, u_4\}$ that $I$ maps `square` to), while $\langle \texttt{\_a\_q x.square(x)} \rangle$ denotes $\{\{u_2\}, \{u_4\}\}$ (i.e., there are two entities that are denoted by the phrase "a square", corresponding to the two entities that $I$ maps `square` to).

These two illustrative examples reveal two general factors that affect the referent $\mathcal{R}$ of $\Phi(r)$. First, every entity in a denotation of $\Phi(r)$ satisfies the restrictor of its generalized quantifier (this is because in this paper we ignore group nouns, such as *committee*). Secondly, the generalized quantifier imposes its own conditions on $\mathcal{R}$, in particular on the cardinality of each set in $\mathcal{R}$ that's a denotation, and for some quantifiers, there is also a constraint on the relationship between a denotation and all the entities in $U$, which is

[2] Each entity $u$ is denoted by a unique constant `u`.

[3] See § 5.2 for details on how $\Phi(r)$ is computed from $r$.

equivalent to a constraint on the cardinality of $\mathcal{R}$ itself. For instance, each set in $\mathcal{R}$ for the referential expression *at least two squares* must have a cardinality of at least two; for *exactly two squares* the cardinality must be equal to 2; and *the two squares* and *both* impose the additional constraint that not only should each denotation have cardinality 2, but also this denotation is *unique* (Russell, 1917) (i.e., $\mathcal{R}$ itself has cardinality 1). Further, as we just mentioned, the referent for *every square* is unique because it's the (unique) maximal set of entities in $\mathcal{M}$ that satisfy the restrictor (again, $\mathcal{R}$ must have cardinality 1).

With this in mind, we obtain a formal definition of the semantics of $\Phi(r)$ for arbitrary $r = \texttt{Qx}.\phi(\texttt{x})$ by first defining a projection of $\mathcal{M}$ onto a smaller model $\sigma(\mathcal{M}, \phi, x)$, which consists of *all and only* those entities $u \in U$ that satisfy $\phi[\texttt{x}/\texttt{u}]$ (i.e., the formula $\phi$ with each occurrence of $\texttt{x}$ substituted with the unique constant $\texttt{u}$ that denotes $u \in U$):

$$\begin{aligned} &\sigma(\mathcal{M}, \phi, \texttt{x}) = (U', V, I') \text{ such that} \\ &U' = \{u \in U_{\mathcal{M}} \mid \mathcal{M} \models \phi[\texttt{x}/\texttt{u}]\} \text{ and} \\ &I' = I_{\mathcal{M}} \downarrow U' \text{ (i.e., } I_{\mathcal{M}} \text{ projected onto } U') \end{aligned} \tag{7}$$

We then define the semantics $\mathcal{R}$ of $\Phi(r)$ for this projected model. As we mentioned, the traditional semantics of a generalized quantifier $\texttt{Q}$ imposes constraints $C_{\texttt{Q}}(R, B)$ on its restrictor $R$ and body $B$. But in our scenario, consisting only of referring expressions, the body $B$ isn't expressed. However, by evaluating the content of the quantifier with respect to the smaller projected model and making $B = U'$ (i.e., the entities in that projected model), the semantics of quantifiers can focus solely on how they constrain the *cardinalities* of these sets to achieve the above-desired effects on referents for $r$. These semantics for each quantifier that we study in our experiments (see Section 5), are defined in column 3 of Table 1. These constraints and the definition of model projection in Eq. 7 yield the following semantics Eq. 8 for the logical form $\Phi(r)$ with the above desired properties:

$$\mathcal{R} = \Phi(r)^{\mathcal{M}} = \langle \texttt{Q x}.\phi \rangle^{\mathcal{M}} = \langle \texttt{Q} \rangle^{\sigma(\mathcal{M}, \phi, \texttt{x})} \tag{8}$$

where $\langle \texttt{Q} \rangle^{\mathcal{M}} = \{R \subseteq U \mid C_{\texttt{Q}}(R, U)\}$ is a referent constructor, utilizing the condition specific to the quantifier (and defined in Table 1 column 3).

To illustrate this symbolic semantics, consider our example model from Eq. 1 and how the referent is computed for a referential expression $r =$ *"the one red square"*, whose logical form is $\langle \texttt{\_the\_1\_q x.red(x)} \wedge \texttt{square(x)} \rangle$. According to Eq. 7, the $\mathcal{M}$-projection consists of the following entities:

$$U' = \{u \in U \mid \mathcal{M} \models \texttt{red(u)} \wedge \texttt{square(u)}\} = \{u_2\} \tag{9}$$

| Quantifier $\mathtt{Q}$ | Surface form | Condition $C_{\mathtt{Q}}(R, B)$ |
|---|---|---|
| `_exactly_n_q` | exactly $n$ | $\lvert R\rvert = n$ |
| `_at_most_n_q` | at most $n$ | $\lvert R\rvert \leq n$ |
| `_at_least_n_q` | at least $n$ | $\lvert R\rvert \geq n$ |
| `_a_q` | a/an | $\lvert R\rvert = 1$ |
| `_every_q` | all/every | $\lvert R\rvert = \lvert B\rvert$ |
| `_the_n_q` | the $n$ | $\lvert R\rvert = \lvert B\rvert \wedge \lvert B\rvert = n$ |
| `_both_q` | both | $\lvert R\rvert = \lvert B\rvert \wedge \lvert B\rvert = 2$ |
| `_all_but_n_q` | all but $n$ | $\lvert R\rvert = \lvert B\rvert - n \wedge \lvert B\rvert \geq n$ |
| `_n_of_the_m_q` | $n$ of the $m$ | $\lvert R\rvert = n \wedge \lvert B\rvert = m$ |

**Table 1:** Generalized quantifiers. The third column shows the condition used by the referent constructor, namely $\langle \mathtt{Q} \rangle^{\mathcal{M}} = \{R \subseteq U \mid C_{\mathtt{Q}}(R, U)\}$. $R$ is a set known as a restrictor and $B$ is a set known as a body in a referential condition. The elements in red highlight the elements of the condition stemming from the Rusellian interpretation of definite descriptions (Russell, 1917).

which in turn leads to the following referent:

$$\begin{aligned} \mathcal{R} &= \langle \mathtt{\_the\_1\_q\ x.red(x) \wedge square(x)} \rangle^{\mathcal{M}} \\ &= \{R \subseteq U' \mid C_{\mathtt{\_the\_1\_q}}(R, U')\} \\ &= \{\{u_2\}\} \end{aligned} \tag{10}$$

If there had been another entity in the domain model, $u_5$ say, such that $\mathtt{red}(\mathtt{u}_5) \wedge \mathtt{square}(\mathtt{u}_5)$, then $\mathcal{R} = \emptyset$. In other words, there is a referential failure because the Russellian uniqueness condition triggered by "the one" (shown in red in Table 1) is violated. By similar symbolic reasoning, the referent for "a blue circle" is $\{\{u_1\}, \{u_3\}\}$ and the referent for "every blue circle" is $\{\{u_1, u_3\}\}$.

During the embodied conversation, a referent to a referring expression $r$ can be provided by the teacher's explicit designation e.g. saying $r$ while pointing to a member of the set $\mathcal{R}$. This can be used to provide a logical formula $\Phi(r)[\mathcal{R}]$ that holds in the domain. For the model in (1), *"the one red square"* while pointing to $u_2$ would produce the following formula:

$$\begin{aligned} \Phi(r)[\mathcal{R}] \equiv\ & \neg(\mathtt{red}(\mathtt{u}_1) \wedge \mathtt{square}(\mathtt{u}_1)) \\ & \wedge (\mathtt{red}(\mathtt{u}_2) \wedge \mathtt{square}(\mathtt{u}_2)) \\ & \wedge \neg(\mathtt{red}(\mathtt{u}_3) \wedge \mathtt{square}(\mathtt{u}_3)) \\ & \wedge \neg(\mathtt{red}(\mathtt{u}_4) \wedge \mathtt{square}(\mathtt{u}_4)) \end{aligned} \tag{11}$$

The negated clauses in Eq. 11 are inferred from the Russellian interpretation (Russell, 1917) for the quantifier `_the_1_q`. By contrast, *"a red square"* and pointing to $u_2$ does not validate any inferences about whether the entities other than $u_2$ are red squares, so it leads to the following formula:

$$\Phi(r)[\mathcal{R}] \equiv \texttt{red}(\texttt{u}_2) \wedge \texttt{square}(\texttt{u}_2) \tag{12}$$

More generally, the teacher's designations of referring expressions $r$ enable the learner to dynamically build a domain theory:

$$\Delta \leftarrow \Delta \cup \{\Phi(r)[\mathcal{R}]\} \tag{13}$$

As we mentioned earlier, we assume that the teacher's utterances and designations are accurate, and furthermore, referential expressions are parsed correctly to their logical forms. So the learner knows that $\mathcal{M} \models \Delta$. In other words, the learner can treat it as monotonic information, which also expands monotonically during the course of the interaction: i.e., no previously acquired facts have to be retracted when acquiring new information.

Such monotonic reasoning is not feasible long-term for two reasons. First, decoding the teacher's message from their signal might be defeasible thanks to both linguistic ambiguity and the defeasible implicatures that get associated with the teacher's utterance in context (Grice, 1975b; Gamut, 1991). Secondly, the knowledge that the teacher imparts may be explicitly defeasible (e.g., they may express a generic statement), thereby requiring non-monotonic reasoning. Supporting symbolic non-monotonic inference, given the teacher's utterance, could be achieved by using e.g. answer-set programming (Lifschitz, 2008) as the inference engine rather than (monotonic) first-order logic.[4] However, we don't opt here for that more complex reasoning for two reasons. Firstly, we don't aim to build from scratch complex domain theories such as those captured in an extensive knowledge base. Instead, we envisage a mode of interaction in which the learner becomes aware of relatively few concepts via a few turns in the conversation, making the need for theory repair and representation revision less urgent (Li et al., 2018; Bundy and Li, 2023). Second, in the particular task that we use to empirically test our model when the teacher and learner encounter a new visual scene, which obviously corresponds to a new model with new entities, the learner retains its beliefs about how to map symbols to entities given their visual features, but the previously built domain theory $\Delta$ must be discarded because it is *not* a (partial) description of the current visual scene (or model) anymore. This

[4] In Barwise and Cooper (1981) witness set is the truth-conditional equivalent of the probabilistic answer set.

reduces the complexity of the theory $\Delta$ that the agent builds over the course of its learning process.

Let's now illustrate the effects of different theories $\Delta$ on probabilistic queries about the domain model. We will consider three separate theories, involving the following three referential expressions, whose semantics relative to the model in Eq. 1 are also given below:

$$r_1 = \textit{“a red square”} \qquad \mathcal{R}_1 = \{\{u_2\}\}$$
$$r_2 = \textit{“the one red square”} \qquad \mathcal{R}_2 = \{\{u_2\}\}$$
$$r_3 = \textit{“the one blue square”} \qquad \mathcal{R}_3 = \{\{u_4\}\}$$

Our three contrasting theories are then defined as follows, where $\phi_1 = \langle \texttt{\_a\_q}\ .\texttt{x}(\texttt{red}(\texttt{x}) \wedge \texttt{square}(\texttt{x}))\rangle[\{u_2\}]$ (and so corresponds to the formulae in Eq. 12), and $\phi_2$ and $\phi_3$ are similarly defined:

- $\Delta_1 = \{\phi_1\}$: built from uttering $r_1$ and the teacher pointing to $u_2$,
- $\Delta_2 = \{\phi_2\}$: built from uttering $r_2$ and the teacher pointing to $u_2$,
- $\Delta_{23} = \{\phi_2, \phi_3\}$: built from uttering $r_2$ and the teacher pointing to $u_2$ and uttering $r_3$ and the teacher pointing to $u_4$.

Table 2 shows the probabilities for queries that result from these different logic theories. Between $\Delta_1$ and $\Delta_2$, there is additional information inferred about the domain, as shown in Eq. 11 vs. Eq. 12. Crucially, this additional monotonic information under WMC affects the probabilities of atoms that are not a part of $\Delta_1$ or $\Delta_2$: e.g., the atom $\texttt{red}(\texttt{u}_1)$ is less likely to be true given $\Delta_2$ compared with its likelihood given $\Delta_1$, because $Pr(\texttt{red}(\texttt{u}_1) \wedge \texttt{square}(\texttt{u}_1)|\Delta_2) = 0$ and this is not the case with respect to $\Delta_1$.

When comparing $\Delta_2$ and $\Delta_{23}$, observe that $\texttt{blue}(\texttt{u}_2)$ and $\texttt{red}(\texttt{u}_4)$ are false thanks to the (classical) logical reasoning from the logical formulae in $\Delta_{23}$. `WMC` yields soft belief changes as well. For example, the neologism "blue" has just been introduced, but one logical consequence of $\Delta_{23}$ is that for all individuals other than $u_4$, the probability that they are both a square and blue is 0. This reduces the likelihood that $u_1$ and $u_3$ are blue for similar reasons to those we mentioned for $\texttt{red}(\texttt{u}_1)$ given $\Delta_2$. In $\Delta_{23}$ we have equivalent evidence for symbols `red` and `blue`, coming from $\phi_2$ ($r_2$=*“the one red square”*) and $\phi_3$ ($r_3$=*“the one blue square”*), respectively leading to the *same* probability assigned for atoms $\texttt{red}(\texttt{u}_1)$, $\texttt{red}(\texttt{u}_3)$, $\texttt{blue}(\texttt{u}_1)$, $\texttt{blue}(\texttt{u}_3)$, while it's different and smaller for symbol `square`, because we have evidence about this symbol from both $\phi_2$ and $\phi_3$. For `MAP` queries, there is no qualitative difference between $\Delta_1$ and $\Delta_2$ because `MAP` sets the truth value of $a$ to true if, and only if, $Pr(a|\Delta) > 0.5$ (according to `EVI`). For $\Delta_{23}$ we infer additional

| Atoms $a$ | $\texttt{EVI}(a, \Delta_1)$ | $\texttt{EVI}(a, \Delta_2)$ | $\texttt{EVI}(a, \Delta_{23})$ | $\texttt{MAP}(\Delta_1)$ | $\texttt{MAP}(\Delta_2)$ | $\texttt{MAP}(\Delta_{23})$ |
|---|---|---|---|---|---|---|
| $\texttt{red}(\texttt{u}_1)$ | .50 | .33 | .40 | 0 | 0 | 0 |
| $\texttt{red}(\texttt{u}_2)$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\texttt{red}(\texttt{u}_3)$ | .50 | .33 | .40 | 0 | 0 | 0 |
| $\texttt{red}(\texttt{u}_4)$ | .50 | .33 | 0 | 0 | 0 | 0 |
| $\texttt{blue}(\texttt{u}_1)$ | – | – | .40 | – | – | 0 |
| $\texttt{blue}(\texttt{u}_2)$ | – | – | 0 | – | – | 0 |
| $\texttt{blue}(\texttt{u}_3)$ | – | – | .40 | – | – | 0 |
| $\texttt{blue}(\texttt{u}_4)$ | – | – | 1 | – | – | 1 |
| $\texttt{circle}(\texttt{u}_1)$ | – | – | – | – | – | – |
| $\texttt{circle}(\texttt{u}_2)$ | – | – | – | – | – | – |
| $\texttt{circle}(\texttt{u}_3)$ | – | – | – | – | – | – |
| $\texttt{circle}(\texttt{u}_4)$ | – | – | – | – | – | – |
| $\texttt{square}(\texttt{u}_1)$ | .50 | .33 | .20 | 0 | 0 | 0 |
| $\texttt{square}(\texttt{u}_2)$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\texttt{square}(\texttt{u}_3)$ | .50 | .33 | .20 | 0 | 0 | 0 |
| $\texttt{square}(\texttt{u}_4)$ | .50 | .33 | 1 | 0 | 0 | 1 |

**Table 2:** Comparison of different probabilistic queries: EVI and MAP for $\Delta_1$,$\Delta_2$ and $\Delta_{23}$. The symbol "$-$" denotes the unawareness about that concept for that particular stage of the interaction. This may change: e.g. upon the teacher conveying $\phi_3$, the vocabulary $V \leftarrow V \cup \{\texttt{blue}\}$ expands, as the agent becomes aware of the concept blue. Thus for $\Delta_{23}$ "$-$" gets replaced with a probability. In MAP queries, 1 means that an atom is in the domain model; 0 means it is not.

atoms to be true in the domain due to the addition of new knowledge from $\phi_3$.

## 3.2 Interactive Symbol Grounding

We now describe how to integrate the evidence stemming from the logical theory $\Delta$, built from the embodied conversation with the (visual) sensory information about the entities and their perceptual similarity. To achieve this, we extend our domain model definition for each entity $u \in U$ to have a corresponding $d$-dimensional feature vector $\mathbf{u} \in \mathbf{U}$ that is extracted from the learner's sensory observation of the environment.[5] For illustration, we amend the toy example domain model from Eq. 1 with 2-dimensional feature vectors for each of its entities, in which the first dimension corresponds to a

[5] Section 5.2 give details about feature extraction

'colour' feature and the second one corresponds to a 'shape' feature:

$$
\begin{aligned}
&\mathcal{M}:\\
&U = \{u_1, u_2, u_3, u_4\}\\
&\mathbf{U} = \{\mathbf{u}_1 = [0.7, 0.2]^\top, \mathbf{u}_2 = [0.1, 0.7]^\top, \mathbf{u}_3 = [0.6, 0.1]^\top, \mathbf{u}_4 = [0.9, 0.8]^\top\}\\
&V = \{\texttt{blue}, \texttt{red}, \texttt{circle}, \texttt{square}\}\\
&I = \{\texttt{blue} : \{u_1, u_3, u_4\}, \texttt{red} : \{u_2\}, \texttt{circle} : \{u_1, u_3\}, \texttt{square} : \{u_2, u_4\}\}
\end{aligned}
\tag{14}
$$

This is for illustrative purposes only: we are using here an extremely small feature space, but the approach we describe for combining logical and sensory information to enhance learning can cope with arbitrarily complex feature spaces, and the feature space used in the experiments in Section 5 is much larger.

Sensory observations must combine with evidence from the embodied conversation to solve the interactive symbol grounding problem. Our grounding model (grounder) uses *prototype networks* $\omega_{\mathcal{S}} : \mathbb{R}^d \mapsto \mathbb{R}^{|V|}$ (Yang et al., 2019; Cano Santín et al., 2020): these offer the means to address a multilabel classification problem with expanding hypothesis space, and they are also designed to support few-shot learning. Here, we show how prototype networks can combine with evidence obtained via the teacher's assertions (i.e., $\Delta$) to ground 1-place predicates (properties), such as shapes and colours. Our model is not limited in principle to grounding 1-place symbols. It could be extended to ground symbols of other arities; e.g., grounding spatial relationships that have binary and ternary arity. But our experiments don't explore the model's capacity to do that in this paper and so to simplify the notation we consider grounding only 1-place symbols.

To ground a symbol $\mathtt{p} \in V$, $\omega_{\mathcal{S}}$ acts as a probabilistic classifier using a sigmoid function: it takes as input a $d$-dimensional feature vector $\mathbf{u} \in \mathbb{R}^d$ and predicts a $|V|$-dimensional *semantic vector* $\hat{\mathbf{y}} \in [0, 1]^{|V|}$, whose entry $\hat{y}_{\mathtt{p}}$ estimates a probability that the atom $\mathtt{p}(\mathtt{u})$ is in $\mathcal{H}_{\mathcal{M}}$.:

$$
\begin{aligned}
\hat{\mathbf{y}} &= \omega_{\mathcal{S}}(\mathbf{u})\\
\hat{y}_{\mathtt{p}} &= \Pr\left(\mathtt{p}(\mathtt{u}) \in \mathcal{H}_{\mathcal{M}} \mid \mathbf{u}, \mathcal{S}\right)\\
\hat{y}_{\mathtt{p}} &= \frac{1}{1 + e^{(-cos(\mathbf{z}_{\mathtt{p}}^- - \mathbf{z}_{\mathtt{p}}^+, f(\mathbf{u}))}}
\end{aligned}
\tag{15}
$$

The sigmoid function uses cosine similarity to compare how visually similar a given entity is compared to positive and negative prototypes, denoted as $\mathbf{z}_{\mathtt{p}}^+$ and $\mathbf{z}_{\mathtt{p}}^-$ respectively. These prototypes are computed using *support*

$\mathcal{S} = \{(\mathbf{u}_i, \mathbf{y}_i)\}_{i=1}^{|\mathcal{S}|}$ of feature vector–semantic vector pairs. This is used to construct positive and negative support for each symbol $\mathtt{p} \in V$, denoted as $\mathcal{S}_{\mathtt{p}}^{+}$ and $\mathcal{S}_{\mathtt{p}}^{-}$, respectively. Whether a pair goes into one of these support sets is decided by the value of $y_{\mathtt{p}}$ as well as the entropy $\mathbb{H}$ of a Bernoulli distribution parametrised by $y_{\mathtt{p}}$. That is, the pair is added to the (positive or negative) support if entropy is smaller than the threshold $\tau$:

$$\begin{aligned} \mathcal{S}_{\mathtt{p}}^{+} &= \{(\mathbf{u}, \mathbf{y}) \in \mathcal{S} \mid y_{\mathtt{p}} > \frac{1}{2} \wedge \mathbb{H}[y_{\mathtt{p}}] \leq \tau\} \\ \mathcal{S}_{\mathtt{p}}^{-} &= \{(\mathbf{u}, \mathbf{y}) \in \mathcal{S} \mid y_{\mathtt{p}} < \frac{1}{2} \wedge \mathbb{H}[y_{\mathtt{p}}] \leq \tau\} \end{aligned} \tag{16}$$

Using these support sets, prototypes are computed by taking the weighted average of the feature vectors of the corresponding support sets, the weight being the likelihood that the relevant entity is (or respectively is not) denoted by the symbol $\mathtt{p}$ (i.e., the values $y_{\mathtt{p}}$ and $(1 - y_{\mathtt{p}})$ respectively).

$$\begin{aligned} \mathbf{z}_{\mathtt{p}}^{+} &= \frac{1}{|\mathcal{S}_{\mathtt{p}}^{+}|} \sum_{(\mathbf{u},\mathbf{y}) \in \mathcal{S}_{\mathtt{p}}^{+}} y_{\mathtt{p}}\, f(\mathbf{u}) \\ \mathbf{z}_{\mathtt{p}}^{-} &= \frac{1}{|\mathcal{S}_{\mathtt{p}}^{-}|} \sum_{(\mathbf{u},\mathbf{y}) \in \mathcal{S}_{\mathtt{p}}^{-}} (1 - y_{\mathtt{p}})\, f(\mathbf{u}) \end{aligned} \tag{17}$$

where $f \colon \mathbb{R}^d \mapsto \mathbb{R}^l$ is an encoder that could be a pre-trained neural network, a small feed-forward neural network or a combination of the two.

The encoder $f$ acts both as a feature extractor and as a way to bound the computation in our approach. As an entity-centric approach, the computational complexity of probabilistic queries depends on the number of entities in the domain: as the number of the entities the agent is aware of grows so do the computational needs, making it intractable for the agent's (perhaps long) lifecycle. To bound the complexity, the knowledge that has been learned in the interaction periodically has to be *integrated* using batch learning techniques (i.e., not during a particular embodied conversation): $f$ is fine-tuned with $\mathcal{S}$ and in turn $\omega_{\mathcal{S}}$ can be used to assign other initial weight values, instead of uniform weights as presented above. Such batch learning is not lossless but does bound the overall complexity of the approach. The fine-tuning of $f$ can be achieved by using binary-cross entropy loss on semantic vectors as an optimization criterion.

$$\mathcal{L}_{BCE} = \sum_{\mathtt{p} \in V} -\hat{y}_{\mathtt{p}} \log\left(y_{\mathtt{p}}\right) - (1 - \hat{y}_{\mathtt{p}}) \log\left(1 - y_{\mathtt{p}}\right) \tag{18}$$

However, exploring empirically the effects of integrating periodic batch learning with the online learning of our model during a 'lifetime' of embodied conversations is beyond the scope of this paper.

Note that if for the lack of evidence $\mathcal{S}_{\mathtt{p}}^{+/-}$ are empty—in other words, there isn't enough evidence (yet) to deem any of the observed feature vectors to be *good-enough* exemplars for positive/negative support—then $\mathbf{z}_{\mathtt{p}}^{+/-}$ defaults to support elements with the largest/smallest Bernoulli entropy $\mathbb{H}$ as the best guess of what exemplars are suitable for these support sets.

The probability distribution $\omega_{\mathcal{S}}$, which assigns to each entity a (binary) probability distribution for each symbol (the symbols being Boolean variables), handles unawareness in the following way. Suppose that the learner observes a new symbol or neologism $\mathtt{p}^*$. $\omega_{\mathcal{S}}$ becomes aware of it by extending the vocabulary $V \leftarrow \{\mathtt{p}^*\} \cup V$ and consequently, each semantic vector is extended with a new entry:

$$\mathbf{y} \leftarrow \texttt{Concat}(\mathbf{y}, 0.5) \tag{19}$$

where `Concat` is a concatenation function, adding the additional value to the prior vector. The value $0.5$ in the (updated) semantic vector represents the (current) complete ignorance about the neologism's denotations. The new prototype network $\omega'_{\mathcal{S}}$ is assigned the same encoder $f$ as the original one $\omega_{\mathcal{S}}$.

$\omega_{\mathcal{S}}$'s prediction for all entities $\mathbf{u} \in \mathbf{U}$ in the domain specifies the weights $\mathbf{w}$ given by the overall grounding model $\Omega_{\mathcal{S}} \colon \mathbb{R}^{d \times |\mathbf{U}|} \mapsto [0,1]^{\mathcal{H}}$:

$$\mathbf{w} = \Omega_{\mathcal{S}}(\mathbf{U}) = [\omega_{\mathcal{S}}(\mathbf{u}_1), \omega_{\mathcal{S}}(\mathbf{u}_2), \ldots, \omega_{\mathcal{S}}(\mathbf{u}_{|\mathbf{U}|})] \tag{20}$$

These weights $\mathbf{w}$ can in turn be used as weights for probabilistic queries. In Rubavicius and Lascarides (2022) we showed how to use $\Delta$ to construct semantic vectors for support $\mathcal{S}'$ using `EVI` queries (see Eq. 5): $y_i = \texttt{EVI}(a_i, \Delta)$ (e.g., see the example values in Table 2). This dynamic change is denoted by the following procedure

$$\mathcal{S}' = \zeta(\mathcal{S}, \Delta) \tag{21}$$

To illustrate how beliefs change, given the evidence $\Delta$ from the embodied conversation combined with sensory observations, we consider the belief changes under the same logic theories $\Delta_1$, $\Delta_2$, and $\Delta_{23}$ as we used in Table 2, combined with the example 2-dimensional vectors for our toy example that are shown in Eq. 14. For this illustrated toy example, we make $f$ identity. But in practice, both the visual features in $\mathbf{U}$ and the extraction is much more complex and hard to reason about—this simple example is designed merely

to anchor one's intuitions. Furthermore, we set the threshold $\tau$ to 0.7. This corresponds to a positive (or negative) exemplar being added to the support set, thereby affecting the vector representation of positive (or negative) prototypes, if the probability that the atom is true is above 0.55 (or for the negative case, below 0.44).

The comparison of the beliefs about grounding using these three different logic theories is given in Table 3. Consider first $\Delta_1$. $\mathbf{u}_2$ is in the positive support vectors for `red` and `square` thanks to $\Delta_1$ (entries in Table 2, column 2 are above 0.55 for these atoms) while $\mathbf{u}_1$ is chosen at random to be assigned to the negative support vector for these concepts, because no exemplars have the probability below the 0.44 margin (see Table 2, column 2): since the possible candidates ($\{\mathbf{u}_1, \mathbf{u}_3, \mathbf{u}_4\}$) all have the same entropy, one is chosen without a loss of generality.

Using these support sets, prototypes for symbols `red` and `square` is $\mathbf{z}^+_{\texttt{red/square}} = \mathbf{u}_2$ and $\mathbf{z}^-_{\texttt{red/square}} = 0.5\mathbf{u}_1$ (0.5 being the probability in Table 2 that $u_1$ is red, and the probability that it is a square). Similarly, computation between these vectors and the feature vector for each entity acts as an activation for the sigmoid in Eq. 15, which yields weights (Table 3, column 6) that show that $u_4$ is more likely to be `red` and `square` than $u_1$ or $u_3$, even though no evidence is provided about that in the embodied conversation. This difference stems entirely from the relevant cosine similarities on the vectors of the entities vs. the positive (and negative) prototypes.

When grounding with $\Delta_2$, the ordering of the likelihoods persists: $u_4$ is more likely to be `red` and `square` than $u_1$ and $u_3$, but the weights are overall lower because the logical theory now has the information stemming from the logical consequences of using "the one" instead of "a".

With domain theory $\Delta_{23}$, compared to the version without sensory observation (Table 2 column 4), integrating sensory information makes a difference, in particular to the (probabilistic) beliefs for $\texttt{red}(\texttt{u}_1)$, $\texttt{red}(\texttt{u}_3)$, $\texttt{blue}(\texttt{u}_1)$ and $\texttt{blue}(\texttt{u}_3)$. In particular, the sensory observations make $\texttt{blue}(\texttt{u}_1)$ and $\texttt{blue}(\texttt{u}_3)$ more probable than $\texttt{red}(\texttt{u}_1)$ and $\texttt{red}(\texttt{u}_3)$, thanks to the visual similarity of $\mathbf{u}_1$ and $\mathbf{u_3}$ to the positive prototypes for blue. These differences lead to a different and more accurate estimated domain model for $\Delta_{23}$ (when compared to using only the symbolic information $\Delta_{23}$ alone): `MAP` query with sensory observations (Table 3 column 7) additionally suggest that $\texttt{blue}(\texttt{u}_1)$ and $\texttt{blue}(\texttt{u}_3)$ holds in the domain, in contrast to the estimated model without sensory observations (Table 2 column 7).

| $\Delta$ | Positive Support Feature Vectors | Negative Support Feature Vectors | Positive Prototypes $\mathbf{z}^+$ | Negative Prototypes $\mathbf{z}^-$ | Weights $\mathbf{w}$ | Model $\mathcal{H}_{\hat{\mathcal{M}}}$ |
|---|---|---|---|---|---|---|
| $\Delta_1$ | $\texttt{red} : \{\mathbf{u}_2\}$<br>$\texttt{blue} : -$<br>$\texttt{circle} : -$<br>$\texttt{square} : \{\mathbf{u}_2\}$ | $\texttt{red} : \{\mathbf{u}_1\}$<br>$\texttt{blue} : -$<br>$\texttt{circle} : -$<br>$\texttt{square} : \{\mathbf{u}_1\}$ | $\mathbf{z}^+_{\texttt{red}} = \begin{bmatrix}.10\\.70\end{bmatrix}$<br>$\mathbf{z}^+_{\texttt{square}} = \begin{bmatrix}.10\\.70\end{bmatrix}$ | $\mathbf{z}^-_{\texttt{red}} = \begin{bmatrix}.35\\.10\end{bmatrix}$<br>$\mathbf{z}^-_{\texttt{square}} = \begin{bmatrix}.35\\.10\end{bmatrix}$ | $\begin{bmatrix}.36, 1., .33, .47\\ -, -, -, -\\ -, -, -, -\\ .36, 1., .33, .47\end{bmatrix}$ | $\{\texttt{red}(u_2)$<br>$\texttt{square}(u_2)\}$ |
| $\Delta_2$ | $\texttt{red} : \{\mathbf{u}_2\}$<br>$\texttt{blue} : -$<br>$\texttt{circle} : -$<br>$\texttt{square} : \{\mathbf{u}_2\}$ | $\texttt{red} : \{\mathbf{u}_1, \mathbf{u_3}, \mathbf{u_4}\}$<br>$\texttt{blue} : -$<br>$\texttt{circle} : -$<br>$\texttt{square} : \{\mathbf{u}_1, \mathbf{u_3}, \mathbf{u_4}\}$ | $\mathbf{z}^+_{\texttt{red}} = \begin{bmatrix}.10\\.70\end{bmatrix}$<br>$\mathbf{z}^+_{\texttt{square}} = \begin{bmatrix}.10\\.70\end{bmatrix}$ | $\mathbf{z}^-_{\texttt{red}} = \begin{bmatrix}.49\\.24\end{bmatrix}$<br>$\mathbf{z}^-_{\texttt{square}} = \begin{bmatrix}.49\\.24\end{bmatrix}$ | $\begin{bmatrix}.27, 1, .26, .31\\ -, -, -, -\\ -, -, -, -\\ .27, 1, .26, .31\end{bmatrix}$ | $\{\texttt{red}(u_2),$<br>$\texttt{square}(u_2)\}$ |
| $\Delta_{23}$ | $\texttt{red} : \{\mathbf{u}_2\}$<br>$\texttt{blue} : \{\mathbf{u}_4\}$<br>$\texttt{circle} : -$<br>$\texttt{square} : \{\mathbf{u}_2, \mathbf{u}_4\}$ | $\texttt{red} : \{\mathbf{u}_1, \mathbf{u_3}, \mathbf{u_4}\}$<br>$\texttt{blue} : \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$<br>$\texttt{circle} : -$<br>$\texttt{square} : \{\mathbf{u}_1, \mathbf{u_3}\}$ | $\mathbf{z}^+_{\texttt{red}} = \begin{bmatrix}.10\\.70\end{bmatrix}$<br>$\mathbf{z}^+_{\texttt{blue}} = \begin{bmatrix}.90\\.80\end{bmatrix}$<br>$\mathbf{z}^+_{\texttt{square}} = \begin{bmatrix}.50\\.75\end{bmatrix}$ | $\mathbf{z}^-_{\texttt{red}} = \begin{bmatrix}.56\\.33\end{bmatrix}$<br>$\mathbf{z}^-_{\texttt{blue}} = \begin{bmatrix}.29\\.29\end{bmatrix}$<br>$\mathbf{z}^-_{\texttt{square}} = \begin{bmatrix}.52\\.12\end{bmatrix}$ | $\begin{bmatrix}.29, 1, .28, 0\\ .41, 0, .42, 1\\ -, -, -, -\\ .20, 1, .19, 1\end{bmatrix}$ | $\{\texttt{red}(u_2),$<br>$\texttt{square}(u_2),$<br>$\texttt{square}(u_4),$<br>$\texttt{blue}(u_1),$<br>$\texttt{blue}(u_3),$<br>$\texttt{blue}(u_4)\}$ |

**Table 3:** Comparison of different support set feature vectors, prototypes, weights, and domain models (MAP query) constructed with different logic theories $\Delta_1$, $\Delta_2$, and $\Delta_{23}$. Semantic vectors $\mathbf{y}$ are omitted as they are equivalent to EVI queries in Table 2. Weights $\mathbf{w}$ are given in the same order for atoms as in Table 2.

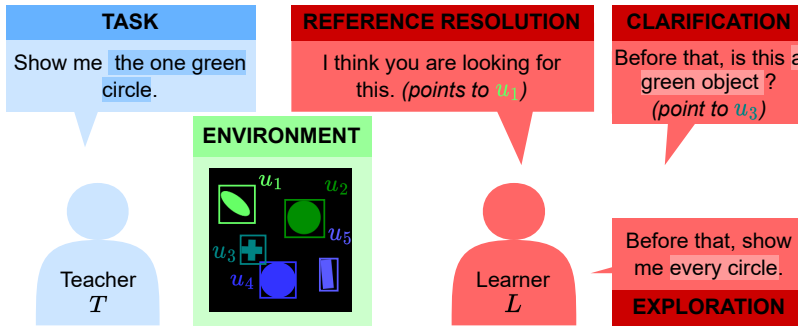



**Figure 4:** A task instance in which learner may refer using its beliefs (currently incorrect) or query the teacher to reduce uncertainty via clarification (asking if an entity is denoted by an expression) or exploration (asking for a referent for an expression).

# 4 Learning a Policy

We now describe the main contribution of this article, which is to extend the method for learning a grounding model from Section 3 with a neuro-symbolic method for learning a policy for acquiring the evidence the learner needs to acquire that grounding model efficiently, through embodied conversation. In effect, this is a method for learning a policy of when (and what) to ask the teacher vs. when (and how) to perform a task in the domain. Like the grounding model, it utilises the logical consequences of quantifiers to reason about what strategies work best. We use this model in our experiments (see Section 5), which address the task of interactive reference resolution, described in Section 4.1. Table 11 in Appendix A gives a summary of symbols and their descriptions from this section.

## 4.1 The Task

Interactive reference resolution involves an embodied conversation between two agents: the learner $L$ and the teacher $T$. We consider an environment of 2D coloured shapes (see Figure 4) based on ShapeWorld (Kuhnle and Copestake, 2017). Each environment instance comes as a pair: an image $\mathbf{X} \in \mathbb{R}^{256\times256}$ that contains entities of various shapes and colours on a solid black background in non-overlapping positions; and a symbolic representation of that image that is much like the domain model $\mathcal{M}$ given in

Eq. 1. Specifically, the symbolic representation features predicate symbols for shapes (domain's vocabulary $V$) in the domain model $\mathcal{M} = (U, V, I)$, where $U$ is a set of entities associated with a feature vector $\mathbf{U}$ that is determined by the sensory observations $\mathbf{X}$ (as illustrated in Eq. 14, for details see Section 5.2), and $I$ is an interpretation function that maps each predicate to a subset of $U$.

The learner $L$ can observe $\mathbf{X}$ and also has perfect object detection: thus for each ShapeWorld instance. $L$ observes $U$ and the corresponding feature vectors $\mathbf{U}$. But $I$ is latent to $L$. Indeed, even the set of possible domain models isn't known to $L$ because it starts its learning process unaware of all colour and shape symbols (its vocabulary $V$ starts as the empty set), and so it doesn't know the domain of $I$, let alone the denotations defined by $I$.

Each interactive reference resolution task $t_r$ is conceptualized as an episode that begins with the teacher $T$ issuing a request to the learner $L$ that it identify a referent for a particular referring expression $r$, such as:

$$T\text{: show me } \underbrace{\text{the one green circle.}}_{r}$$

$L$ observes the logical form $\Phi(r)$ of $r$: see Section 3.1 for the symbolic logic of $\Phi(r)$ and Section 5.2 for how $L$ constructs $\Phi(r)$, even when $r$ contains a neologism.

$L$'s objective is to accurately point to a referent of $r$. $L$'s estimated referent(s) of $r$ is $\Phi(r)^{\hat{\mathcal{M}}}$ (Eq. 8), where $\hat{\mathcal{M}}$ is the learner's estimated domain model (Eq. 6). In this interaction, $L$ faces a choice: either point to a currently estimated referent; or utter a query to $T$ so as to use her response to minimize the uncertainty about what to point to. Queries come at a cost, which incentivises $L$ to minimise $T$'s effort.

If $L$ decides its next action is to point at an estimated referent (which may involve pointing to more than one entity, e.g., for expressions such as "every circle" or "both squares"), then it executes the following (the estimated referent being given in red):

$$L\text{: Here it is } (\textit{points to } \underbrace{\{\{u_1\}\}}_{\Phi(r)^{\hat{\mathcal{M}}}}).$$

By performing this action, $L$ receives a reward: $R = 1$ if the referent is correct, $R = -1$ if it isn't. This action terminates the episode. If, on the other hand, $L$ chooses to utter a query, then the episode continues, $L$ incurs the query's cost and uses $T$'s response to compute updated beliefs. A sequence of interactive resolution tasks $\mathcal{T} = t_{r_1}, t_{r_2}, \ldots, t_{r_{|\mathcal{T}|}}$ over the *same image* $\mathbf{X}$

constitutes an *embodied conversation* $\mathcal{C} = (\mathbf{X}, \mathcal{T})$. A sequence of embodied conversations (i.e., one conversation per image) constitutes a *dataset* $\mathcal{D} = \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_{|\mathcal{D}|}$.

As we've stated, $L$ can query the teacher to improve its beliefs. We consider two types of *queries* $q \in \mathcal{Q}$. The first type is an *exploration query*, which allows $L$ to find out more about the symbols that feature in $r$. E.g., in our running example:

$$L\text{: Before that, show me } \underbrace{\text{every circle}}_{r_i}\text{.}$$

$r_i$ must include a subset of $r$'s non-logical symbols plus a quantifier that's restricted to guarantee a valid reference. E.g., where $r$ is "the one green circle", $r_i$ could be "the one green circle", "a circle", "one green object", "a green circle", "every circle", and so on. But "the one circle" and "two green circles" are forbidden: the former carries a Russellian uniqueness condition (there is exactly one circle) that might be false, and the latter violates the uniqueness condition entailed by $r$ (see column 3 of Table 1). The above query results in a response where $T$ designates a referent of $r_i$, like this:

$$T\text{: Here it is. (points to } \underbrace{\{\{u_2, u_4\}\}}_{\Phi(r_i)^{\mathcal{M}}}\text{)}$$

The second category of queries is *clarification queries*: *yes/no*-questions about a particular entity, e.g.:

$$L\text{: Is this } \underbrace{\text{a circle}}_{r_j}\text{? (points to } \underbrace{u_3}_{u \in U}\text{)}$$

$r_j$ is similar to $r_i$, except the quantifier is "a", "an" or "the one" to ensure $r_j$'s denotations are single entities. $T$'s response is:

$$T\text{: } \begin{cases} \text{Yes. (if } \{u\} \in \Phi(r_j)^{\mathcal{M}}\text{)} \\ \text{No. (otherwise)} \end{cases}$$

Each query has an inherent cost, $\texttt{Cost}\colon \mathcal{Q} \mapsto \mathbb{R}_{>0}$. As we'll see in Section 4.2, this will allow us to balance $T$'s effort to find and show the referent against the value of $L$ knowing $T$'s answer. We approximate the inherent cost of answering the query using two quantities: the number of entities $Ent(q)$ in the referent of the query, which is determined by the head quantifier for a referential expression and approximates the pointing effort; and the number

of symbols $Sym(q)$ in the logical form of $q$, which approximates the effort to search for referents. These quantities are weighted by unit pointing $w_{point}$ and unit symbol reference $w_{ref}$, resulting to the overall cost defined as follows:

$$\texttt{Cost}(q) = w_{point}Ent(q) + w_{ref}Sym(q) \tag{22}$$

To illustrate, consider $q$ = *"before that show the one red square"*. The cost of $q$ would be $w_{point} + 2w_{ref}$: $q$ features two symbols and its answer must point to one entity. For some of the quantifiers (e.g., `_every_q`) the number of entities in a referent is not known in advance: it could be anything from 1 to $|U|$. To approximate this, we use an average number of entities assuming ignorance about reference (e.g., $\frac{|U|-1}{2}$ for `_every_q`).

## 4.2 The Neuro-symbolic Learner

We now present a neuro-symbolic learner that aims to solve interactive reference resolution tasks by learning a policy from evidence, the evidence being the observed costs or rewards of the actions it has experienced so far in various embodied conversations for the purpose of doing reference resolution. Our model is novel in that learning such a policy utilizes reasoning about the logical consequences of possible responses to its queries. Specifically, as we'll see shortly, the learner distinguishes the expected reward of a query such as *show me every circle* from *show me a circle*. It also copes with the discovery of, and adaptation to, unforeseen concepts.

As is standard in models of rational behaviour, our learner's decisions aim to maximise the expected reward. This is determined by its epistemic state, which captures its expectations of what outcome (and hence what reward) it is likely to achieve from its actions. We denote the learner's current (epistemic) state as a tuple $s = (\Omega_{\mathcal{S}}, \Delta, \mathbf{U}) \in S$, consisting of the current grounding model $\Omega_{\mathcal{S}}$ with support $\mathcal{S}$, the current domain theory $\Delta$ (as constructed from the teacher's assertions, described in Section 3.2) and the set of visual feature vectors $\mathbf{U}$.

On receiving the teacher's utterance (e.g., the teacher's response to the learner's query $q$) with logical form $\phi$, the state is updated in the following manner:

$$\texttt{Update}(s, \phi) = \texttt{Update}((\Omega_{\mathcal{S}}, \Delta, \mathbf{U}), \phi) = (\Omega_{\zeta(\mathcal{S}, \Delta \cup \{\phi\})}, \Delta \cup \{\phi\}, \mathbf{U}) \tag{23}$$

When faced with the task of finding a referent of $r$, the learner can either ask a query $q \in \mathcal{Q}$ (subject to the above constraints on the quantifier and

non-logical symbols in $q$), or it can perform the action `ref` of designating a referent based on $\hat{\mathcal{M}}$. Thus the action space $A = \{\texttt{ref}\} \cup \mathcal{Q}$.

As we mentioned earlier, for $\texttt{ref} \in A$, the reward is $R = 1$ if the correct referent is found and $R = -1$ otherwise. The (immediate) reward of asking query $q \in \mathcal{Q}$ approximates the trade-off between the information that the learner gains from the teacher's response—in other words, how much the new epistemic state improves its chances of correctly identifying the referent—and the teacher's effort in responding, as given by the `Cost` function in Eq. 22.

We use a *preference* function $H\colon S \mapsto \mathbb{R}$ (Sutton and Barto, 1998) to quantify the favourability of being in $s$. In words, $H$ is determined by the learner's confidence that it will succeed in performing the task, given $\hat{\mathcal{M}}$.[6] It is defined in terms of the weights $\mathbf{w}$ (Eq. 20) within the learner's (current) grounding model $\Omega_{\mathcal{S}}$. Because of the expanding domain model (the size of $\mathbf{w}$ expands every time the learner discovers a new concept via the teacher's neologisms), functional approximation methods requiring a fixed size input are not immediately suitable in our scenario. To cope with this, we use a statistic extractor function $h : [0,1]^{|\mathcal{H}|} \mapsto \mathbb{R}^m$ to extract $m$ summary statistics from a distribution defined by $\mathbf{w}$. Using these extracted features, $H$ is computed by a linear approximation parameterized by $\theta = \{\mathbf{v}, b\}$:

$$H(s) = \mathbf{v}^\top h(\mathbf{w}) + b = \mathbf{v}^\top h(\Omega_{\mathcal{S}}(\mathbf{U})) + b \tag{24}$$

Using $H$ and `Cost`, the reward for asking $q$ is defined as follows, where $\texttt{ANS}(q)$ is the logical form of the answer that the teacher gives to $q$:

$$R(q) = \texttt{Softmax}\left(\frac{H(\texttt{Update}(s, \texttt{ANS}(q))) - H(s)}{\texttt{Cost}(q)}\right) - 1 \tag{25}$$

In words, $R(q)$ depends on the difference in preferences $H$ of the updated epistemic state that's triggered by the teacher's answer to query $q$ (compared with the prior belief state) and the query's cost $\texttt{COST}(q)$: it effectively measures the gain in information per unit cost. This quantity is normalised between the queries using the `Softmax` function over possible queries $\mathcal{Q}$.

Note that, like the reward for `ref`, the learner will not observe this reward until the teacher's response is observed—the learner does not know what answer $\texttt{ANS}(q)$ the teacher will provide until after the learner has actually asked $q$. So the learner must base its decisions on *expected* rewards, given its current epistemic state $s$. We'll shortly define how the agent computes these expected rewards for both `ref` and for a query $q$. But overall, it means that, in

[6] Given this metric one could view $H$ has measured the informativeness of $s$, but we stick with the term preference to match the reinforcement learning literature.

the usual way, the learner must base its decision on an action-value function $Q(a \mid s)$ that captures these expected rewards. Specifically, the learner can choose a (greedy) $a^* \in A$ as follows:

$$a^* = \arg\max_{a \in A} Q(a \mid s) \tag{26}$$

Now let's define these expected rewards $Q(a \mid s)$. The expected reward $Q(\texttt{ref} \mid s)$ is defined as follows, where $Pr(\texttt{ref}|s)$ is the probability of successful task execution, given the epistemic state $s$:

$$Q(\texttt{ref} \mid s) \triangleq 2Pr(\texttt{ref} \mid s) - 1 \tag{27}$$

The reward for correct and incorrect execution is 1 and $-1$ respectively, so the computation simplifies to the one given in Eq. 27. Where $r$ is the referring expression for which the teacher requested a referent ("show me $r$"), $\Pr(\texttt{ref} \mid s)$ is computed using EVI, where $\Phi(r)[\Phi(r)^{\hat{\mathcal{M}}}]$ is a logical form that's constructed from the estimated referent $\Phi(r)^{\mathcal{M}}$:

$$\Pr(\texttt{ref} \mid s) = \texttt{EVI}(\Phi(r)[\Phi(r)^{\hat{\mathcal{M}}}], \Delta) \tag{28}$$

where For instance if $r$=*"a red square"* and the estimated domain model computes the referent to be $u_1$, then we are evaluating the probability of $\texttt{red}(\texttt{u}_1) \wedge \texttt{square}(\texttt{u}_1)$ using the methods described in Section 3.

The expected reward $Q(q|s)$ effectively replaces the numerator in Eq. 25, which captures the difference in preference afforded by the *actual* answer to $q$, with the established concept of *Value of Information* (VOI)—that is, the expected difference in preference, as determined by marginalising over all the possible answers to $q$ (Howard, 1966):

$$\texttt{VOI}(q \mid s) = \mathbb{E}_{\phi \sim \texttt{ANS}(q)}[H(\texttt{UPDATE}(s, \phi))] - H(s) \tag{29}$$

More specifically, the expectation in Eq. 29 is computed by marginalising over every possible answer $\phi$ to the query $q$, consistent with the learner's current state $s$ (Eq. 30), with the probability that the teacher answers with $\phi$ computed by marginalising over all the models $\mathcal{M}$ that are consistent with $\phi$,

with that probability computed via WMC (Eq. 32), as defined in Eq. 4:

$$\mathbb{E}_{\phi\sim\texttt{ANS}(q)}[H(\texttt{UPDATE}(s,\phi))] = \tag{30}$$

$$= \sum_{\phi} \Pr\left(\texttt{ANS}(q) = \phi \mid s\right) H(\texttt{UPDATE}(s,\phi))$$

$$= \sum_{\phi} \Pr_{\Omega_{\mathcal{S}}}(\phi) H((\Omega_{\zeta(\mathcal{S},\phi)}, \Delta \cup \{\phi\}, \mathbf{U})) \tag{31}$$

$$= \sum_{\phi} \texttt{WMC}(\phi, \Omega_{\mathcal{S}}(\mathbf{U})) H((\Omega_{\zeta(\mathcal{S},\phi)}, \Delta \cup \{\phi\}, \mathbf{U})) \tag{32}$$

Thus $Q(q \mid s)$ works out to be like $R(q)$ except VOI is used for information gain:

$$Q(q \mid s) \triangleq \texttt{Softmax}\left(\frac{\texttt{VOI}(q \mid s)}{\texttt{Cost}(q)}\right) - 1 \tag{33}$$

$Q$ parameters $\theta = \{\mathbf{v}, b\}$ are learned by semi-gradient SARSA (Rummery and Niranjan, 1994), the particular instantiation of which is outlined in Algorithm 1.

We have modelled the problem of decision-making for reference resolution as a single-shot decision-making problem, making the learner 'myopic' (in other words, when deciding on its next action, the learner does not speculate about the effects that the teacher's response to a query might have on the expected value of follow-up queries). We've done this to make the overall problem tractable. Treating this as a single-shot decision problem contrasts with, for instance, partially observable Markov Decision Processes (POMDPs), which support sequential decision-making, trading off immediate rewards against longer-term returns (Mykel J. Kochenderfer and Wray, 2022). But as well as the vastly increased complexity that comes with solving a sequential decision problem, our problem cannot be modelled, even in principle, by a POMDP. This is because POMDPs don't support a changing hypothesis space of possible states and actions, and provide no reasoning about how to utilise one's experience and learning so far on the discovery of an unforeseen possibility. At any rate, while decision-making in our model is myopic, it does not exclude the option of asking a sequence of queries before acting in the environment (see examples of such cases that are attested in our experiments in Section 5.3).

Another interesting feature of our decision-making problem is that the reward function is non-stationary when learning dialogue strategy (Eq. 25 depends on preference function, and in turn $\theta$). This is done so that the reward function itself would reflect the value of changing the state to a more

favourable state. Such a choice makes the policy learning procedure more stochastic, but during evaluation (as done in Section 5), the reward function is stationary and is much closer to the traditional decision-making problem.

**Algorithm 1** semi-gradient SARSA for dialogue strategy learning

**Require:** learning rate $\alpha$, discount factor $\gamma$, epsilon $\epsilon$, dataset $\mathcal{D}$, action-value function $Q$ parameterized by $\theta$, grounding model $\Omega$.

**for** TASKS, $\mathbf{U}$ in $\mathcal{D}$ **do** ▷ iterate over the dataset
  $\Delta \leftarrow$ THEORY() ▷ initial theory
  $s \leftarrow (\Omega, \Delta, \mathbf{U})$ ▷ initial state
  **for** TASK in TASKS **do** ▷ process task instance
    $\mathcal{Q} \leftarrow$ QUERIES(TASK) ▷ queries as defined in § 4.1
    $A \leftarrow \{ref\} \cup \mathcal{Q}$ ▷ action space
    **while** $True$ **do** ▷ SARSA training loop
      $a \leftarrow \arg\max_{a \in A} Q(s; A)$ ▷ greedy action
      $R, s', Done \leftarrow$ ACT($a$) ▷ act in the environment
      **if** $Done$ **then**
        $\delta \leftarrow R - Q(s, a)$
        $\theta \leftarrow \theta + \alpha \delta \nabla_\theta Q(s, a)$
        **break**
      **else**
        $a' \leftarrow \begin{cases} \arg\max_{a \in A} Q(s'; A) \text{ with prob. } 1 - \epsilon \\ \text{CHOOSE}(A) \text{ otherwise} \end{cases}$
        $\delta \leftarrow R + \gamma Q(s', a') - Q(s, a)$
        $\theta \leftarrow \theta + \alpha \delta \nabla_\theta Q(s, a)$
      $s \leftarrow s'$

# 5 Experiments

We now describe the experiments we use to evaluate the effects of utilizing the logical consequences of quantifiers for both decision-making and grounding when the task to be mastered is interactive reference resolution.[7]

[7]The code and data for these experiments can be found in `https://github.com/ipab-rad/dialogue-strategies`.

| Dataset | # Conversations | # Tasks | # Task / # Conversations |
|---|---|---|---|
| $\mathcal{D}_{train}$ | 128 | 496 | 3.9 |
| $\mathcal{D}_{test}$ | 32 | 51 | 1.6 |
| $\mathcal{D}^*_{test}$ | 32 | 128 | 4 |

**Table 4:** Number of conversations, number of tasks, and average number of tasks per conversation in each experiment datasets.

## 5.1 Experimental Setup

Our experiments use three ShapeWorld datasets: $\mathcal{D}_{train}$, $\mathcal{D}_{test}$, and $\mathcal{D}^*_{test}$. Each of these datasets consists of a set of ShapeWorld images that contain 5-7 coloured shapes in varying orientations and hues. These objects can be described using the 7 shapes (`square`, `rectangle`, `circle`, `ellipse`, `triangle`, `cross`, `pentagon`) and 7 colours (`red`, `blue`, `green`, `yellow`, `magenta`, `cyan`, `grey`). The orientation of entities in each ShapeWorld instance and the shades of the colour categories (which are not mutually exclusive) can both vary within a particular ShapeWorld instance $\mathbf{X}$ and also across the different ShapeWorld instances that $L$ and $T$ will encounter during learning. For each ShapeWorld image $\mathbf{X}$ in the dataset $\mathcal{D}$, there is a sequence of tasks $\mathcal{T}$, where each $t \in \mathcal{T}$ is an instruction to identify a referent of a referring expression $r$ ($r$ is chosen so that it has a non-empty denotation in $\mathbf{X}$). As we mentioned earlier, the verbal exchanges that ensue while performing the sequence of tasks in $\mathcal{T}$ constitute a conversation $\mathcal{C}$. The number of entities in a correct referent varies in the tasks from 1 to 5, with the average number being 1.4. Table 4 gives details of the number of conversations (equivalent to the number of images) and the number of tasks in each data set.

$\mathcal{D}_{test}$ and $\mathcal{D}^*_{test}$ consist of the same images, but their tasks are different. The symbols used in the referring expressions $r$ for tasks $\mathcal{T}$ in $\mathcal{D}_{train}$ and $\mathcal{D}_{test}$ are *restricted*: they feature only 10 of the 14 possible symbols: `square`, `rectangle`, `triangle`, `cross`, `ellipse`, `red`, `blue`, `green`, `cyan`, `grey`. In contrast, $\mathcal{T}$ in $\mathcal{D}^*_{test}$ features the additional 4 symbols: `circle`, `pentagon`, `yellow`, `magenta`. This enables us to showcase in our experiments the extent to which the learner copes with unforeseen possibilities at test time, having experienced different unforeseen possibilities during training. We should stress, however, that for both test data sets, $\mathcal{D}_{test}$ and $\mathcal{D}^*_{test}$, the agent retains its learned policy from training, but we re-set the agent so that it starts the test phase unaware of all shape and colour words and completely ignorant about their denotations and what they look like: we're testing in both test

data sets how well the learned policy copes with overcoming unawareness so as to master interactive reference resolution.

Table 5 outlines the number of occurrences of each symbol in the underlying truth domain model $\mathcal{M}$ (a.k.a. the maximum number of exemplars that could be observed if the learner $L$ knows $\mathcal{M}$) and the number of symbols mentioned in each task $t_r$ (not all entities get covered in all conversations).

The tasks in each conversation have overlapping concepts (e.g., *"show me the one red square"* and *"show me a red circle"*), but tasks within a single conversation are not directly reducible to each other without sensory observations (e.g. the sequence of tasks $\mathcal{T}$ doesn't ever include both *"show me a square"* and *"show me the one square"*). The learner maintains the same domain theory $\Delta$ between the tasks in the same conversation, for $\Delta$ is building up a partial, but ever more specific, symbolic description of the same image, or equivalently the same model $\mathcal{M}$. But when the image changes, the learner must estimate a new model. Accordingly, the learner retains what it's learned so far about the mapping from symbols to visual features (ie., the support $\mathcal{S}$, from which visual prototypes $\mathbf{z}^{+/-}$ and grounding model $\Omega_{\mathcal{S}}$ are computed). But the conversation is no longer about the same entities, and so the learner must reset $\Delta$ to $\emptyset$ and re-set $\hat{M}$ via its (accurate) object detection, their associated visual features and hence, via $\Omega_{\mathcal{S}}$, the prior probabilities of which of these (new) entities in the new image are denoted by which symbols.

As argued in Section 3.2, this support $\mathcal{S}$ could periodically be integrated into the encoder $f$, to bound the size of the support. However such integration is out of the scope of this article. Instead, we just maintain support from all conversations, and the grounding model $\Omega_{\mathcal{S}}$ makes decisions based on it.[8]

In the training phase, the learner optimizes $Q$ parameterised by $\theta$ using semi-gradient SARSA, as given in Algorithm 1, by being exposed to $\mathcal{D}_{train}$. The learner starts unaware of *all* shapes and colour symbols and their mappings to referents, and has to acquire both by learning an effective dialogue policy for learning a grounding model $\Omega_{\mathcal{S}}$ (see §3.2).

In the testing (evaluation) phase, the learner uses $Q$ parameterized by $\theta$ that it learned in training, but as we mentioned earlier: it starts the test phase completely unaware of shape and colour terms again: we do this by re-setting its vocabulary $V$ to $\emptyset$ (i.e., the agent is made unaware of all the colour and shape symbols), and its support $\mathcal{S}$ and grounder $\Omega_S$ that it acquired while training gets dropped as well (so that, on encountering a neologism, it starts out completely ignorant about what the denotations look like). Thus during the test phase, the learner is tasked with learning the grounding model again

[8] Situations where entities are added to the domain is deemed as a new conversation and the logical theory $\Delta$ would be discarded in such a situation while support $\mathcal{S}$ persists.

| Symbols | $\mathcal{D}_{train}$ | | $\mathcal{D}_{test}$ | | $\mathcal{D}^*_{test}$ | |
|---|---|---|---|---|---|---|
| | in $\mathcal{M}$ | in $t_r$ | in $\mathcal{M}$ | in $t_r$ | in $\mathcal{M}$ | in $t_r$ |
| `blue` | 125 | 42 | 41 | 14 | 41 | 16 |
| `red` | 114 | 38 | 24 | 9 | 24 | 13 |
| `green` | 137 | 51 | 31 | 10 | 31 | 12 |
| `yellow` | 126 | - | 42 | - | 42 | 17 |
| `cyan` | 119 | 41 | 43 | 11 | 43 | 13 |
| `magenta` | 147 | - | 29 | - | 29 | 9 |
| `grey` | 138 | 27 | 30 | 15 | 30 | 18 |
| Average in category | 129 | 28 | 34 | 8 | 34 | 14 |
| Total in category | 906 | 199 | 240 | 59 | 240 | 98 |
| `circle` | 123 | - | 35 | - | 35 | 14 |
| `cross` | 116 | 45 | 33 | 14 | 33 | 21 |
| `pentagon` | 127 | - | 30 | - | 30 | 12 |
| `rectangle` | 152 | 56 | 38 | 9 | 38 | 11 |
| `square` | 87 | 37 | 25 | 6 | 25 | 7 |
| `triangle` | 134 | 37 | 43 | 13 | 43 | 18 |
| `ellipse` | 167 | 36 | 43 | 15 | 43 | 18 |
| Average in category | 129 | 30 | 34 | 8 | 34 | 14 |
| Total in category | 906 | 211 | 240 | 57 | 240 | 97 |
| Average symbols | 129 | 29 | 34 | 8 | 34 | 14 |
| Total symbols | 1812 | 410 | 480 | 116 | 480 | 195 |

**Table 5:** Number or symbols in each dataset used in the experiments, by category (colour and shape) together with the total count and the average number of symbols per category. Columns "in $\mathcal{M}$ " show the number of symbol occurrences in the ground-truth domain model while Columns "in $t_r$" shows the number of times a symbol occurs in the tasks. $-$ indicates that the relevant symbol was not introduced during the learning for the entire dataset $\mathcal{D}$.

'from scratch', starting in a position of unawareness, but using it's learned policy (of when to query and when to act in the environment). The learner is exposed to $\mathcal{D}_{test}$ featuring the same symbols as those in $\mathcal{D}_{train}$, and (separately) to $\mathcal{D}^{*}_{test}$, which includes symbols not in $\mathcal{D}_{train}$.

To evaluate the effects of reasoning with the logical consequences of quantifiers on learning a dialogue strategy that aims for accurate grounding, and on learning the grounder itself, we consider the following types of learners:

- Semantic learner $L_{sem}$ performs grounding utilising the logical consequences of quantifiers to construct (noisy) support, and makes decisions on whether to query the teacher that are based on the expected value of information (Eq. 29) that takes these consequences into account as well, thereby making the expected value of a query sensitive to its quantifier (i.e., the quantifier affects both `VOI` in Eq. 29 and $Ent(q)$ in Eq. 22).
- Base learner $L_{base}$ is just like $L_{sem}$ except it doesn't have any active learning: that is, it lacks entirely the capacity to query $T$. It uses only $T$'s yes/no response to the action `ref` to acquire its grounder.
- The 'simple' learner $L_{exists}$ performs simplified semantic analyses, in which each quantifier in $\Phi(r)$ is replaced with the existential $\exists$. That means that when deliberating over the expected value of its queries, they are all the same: i.e., it always amounts to the trade-off between the value of information and the cost of using an existential (see Eqs. 22 and 29). Likewise, the formulae added to $\Delta$ during grounding is that of the existential referring expression, as contrasted with those for other quantifiers, regardless of the quantifier that the teacher used. So overall, the interpretations and potential benefits that $L_{exists}$ ascribes to the referential expressions "both circles", "two circles", "a circle" and "every circle" are all the same, when performing grounding and when learning a dialogue strategy.

We further consider the interaction between using the logical consequences of quantifiers, or not, on each of the two learning tasks—to learn a grounder, and to learn a dialogue strategy—by conducting an ablation study on two mixture learners:

- $L_{mix_1}$ makes decisions like $L_{sem}$ but performs grounding like $L_{exists}$: i.e., its strategy discriminates among the meanings of quantifiers but its grounding does not.

- $L_{mix2}$ makes decisions like $L_{exists}$ but performs grounding like $L_{sem}$: i.e., its grounding discriminates among the meanings of quantifiers but its strategy does not.

We compare the different learners with the following metrics:

- The *cumulative reward* received in $\mathcal{D}_{test}$ and $\mathcal{D}_{test*}$, showing the value of the policy used. Each learner aims to maximise this metric in its decision-making, choosing an action with the maximum expected value (Eq. 26), with expected values defined in Eqs. 27–33 (where for the simple learner $L_{exists}$ these are computed off the existential regardless of the quantifier in the utterances). Note that the base learner $L_{base}$ doesn't attempt to resolve dilemmas between asking vs. doing, because querying the teacher isn't an option for it.
- Macro average *F1 score* on the learner's chosen referent vs. the true referent for each task.

We do not measure F1 scores on $\hat{\mathcal{M}}$ because there are frequently symbols with empty denotations in the given $\mathcal{M}$, in which case F1 is 0. Those symbols with empty denotations vary across the models, so it is not informative or meaningful to measure average F1 scores over all the domain models.

## 5.2 Implementation Details

To extract feature vectors $\mathbf{U}$ from the image $\mathbf{X}$, we use bounding boxes from the generation process to localize each entity in the visual scene and DENSENET161 (Huang et al., 2017) for feature extraction.

To process natural language, and in particular, to parse referential expressions to their logical form, we use the same pipeline as used in Rubavicius and Lascarides (2022). It consists of the English Resource grammar (ERG) (Copestake and Flickinger, 2000) and ACE[9] parser to obtain a minimal recursion semantics (MRS) representation (Copestake et al., 1997), which is further processed using UTOOL (Koller and Thater, 2005) to remove underspecification, and customised munging rules then yield the logical forms of referential expressions in the format needed for our reasoning component. ERG is a wide-coverage hand-crafted grammar with a formal semantic component, which handles unknown words via (estimated) part-of-speech tags. It defines a procedure for constructing the predicate symbol for the unknown

[9] `http://sweaglesw.org/linguistics/ace/`

word from its orthography, with the symbol's arity determined by the syntactic parse that utilizes the unknown word's part-of-speech. This ensures well-formed logical forms $\Phi(r)$ for expressions $r$ with neologisms.

To compute probabilistic queries (`WMC`, `EVI` and `MAP`), as well as to estimate the value of information (`VOI`) for decision-making, we utilise Probabilistic Sentential Decision Diagrams (PSDDs) (Kisa et al., 2014).[10] The logical forms of referential expressions are in general in predicate logic, but to perform inference, they are propositionalized to propositional logic and converted to conjunctive normal form. This conversion and inference are in principle NP-hard (Valiant, 1979), but in our experiments and in practice reasoning over a small bounded set of formulae can be done in real-time.

Table 6 outlines the hyperparameters and their values that are used in our experiments for learning dialogue strategies and grounding models. We used standard values (e.g. discount factor $\eta = 0.98$) rather than performing tuning, to avoid overfitting within the low-data regime. For the `Cost` (Eq. 22), we used the same for both symbol usage $c_s$ and entity designation $c_e$, scaled down to match the order of magnitude of the value of information observed. These values are set to encourage exploration initial exploration. For example, if for simplicity we ignore the value of the information term, the clarification question "Show me a one red square" has an overall cost of 0.3 which the agent may choose to incur unless it belief about the the denotation of "a red square" is above 0.65.

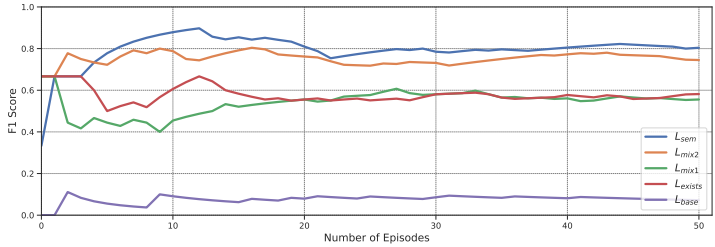
## 5.3 Results

Since we are interested in the rate of learning during the test phase, and not merely what the agent knows or achieves at the end, we present the results of our experiments as graphs, with the $x$-axis being the sequence of tasks and the $y$-axis being the cumulative reward or F1-score. Figures 5a and 5b record the change to F1 scores on the learner's chosen referent over $\mathcal{D}_{test}$ and $\mathcal{D}^*_{test}$ respectively (recall that the learner starts the test phase with no information about grounding). They show that $L_{sem}$ grounds more efficiently than the other types of learners: F1 scores for the final `ref` actions benefit from the entire teacher-learner interaction and are significantly higher ($t$-test, $p < 0.05$ against each other type of learner).

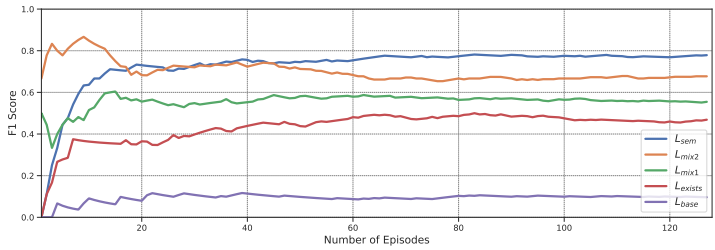
Figures 5c and 5d record the changes to the cumulative reward over $\mathcal{D}_{test}$ and $\mathcal{D}^*_{test}$. The curves take a downward trajectory because all queries receive a negative reward. $L_{sem}$ performs better than the other learners ($t$-test, $p <$

[10] Implemented in the PyPSDD package: `https://github.com/art-ai/pypsdd`

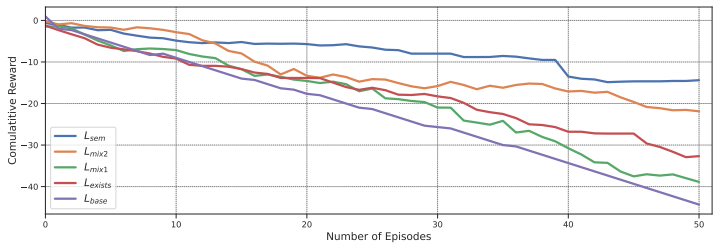
[11] $h$ computes mean, variance, standard deviation, skew, kurtosis, and entropy of $\hat{\mathbf{c}}$



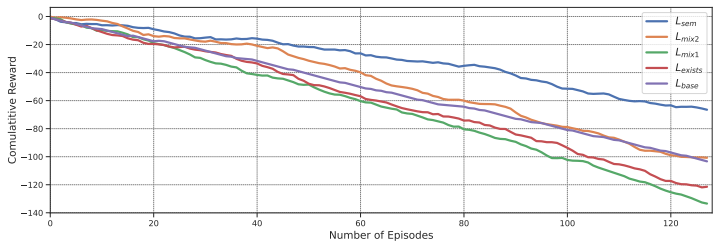

**(a)** F1 score on $\mathcal{D}_{test}$




**(b)** F1 score on $\mathcal{D}^*_{test}$




**(c)** Cumulative Reward on $\mathcal{D}_{test}$




**(d)** Cumulative Reward on $\mathcal{D}^*_{test}$

**Figure 5:** The performance measured by the cumulative reward and macro average F1 of the learners on different test sets: $\mathcal{D}_{test}$ and $\mathcal{D}^*_{test}$ evaluated over 10 random initialisation of sequencing $\mathcal{D}_{train}$.

| Parameter | Symbol | Value |
|---|---|---|
| Cost function | `Cost` | |
| - symbol usage costs | $c_s$ | 0.1 |
| - entity designation cost | $c_e$ | 0.1 |
| Grounding model | $\omega$ | |
| - input size | $m$ | 1000 |
| - embedding size | $l$ | 5 |
| - threshold | $\tau$ | 0.7 |
| Action-value function | $Q$ | |
| - learning rate | $\alpha$ | 0.1 |
| - discount factor | $\eta$ | 0.98 |
| - epsilon | $\epsilon$ | 0.25 |
| - number of features for $h^{11}$ | $v$ | 6 |

**Table 6:** Hyperparameter values used in the experiments.

0.01 against each other learner); flattening curves show that toward the final episodes, learners ask fewer queries while performing successful `ref` actions.

In Figure 5d, $L_{base}$ performs better than $L_{exists}$ and $L_{mix1}$ because it does not issue queries and so doesn't receive the negative rewards that the other learners do. However, Figures 5a and 5b show that $L_{base}$ has the worst performance among all learners on selecting correct referents, given the sample sizes. The fact that it fails to master the task is hardly surprising. The only evidence it receives is the binary reward signal "yes/no" from the teacher, and it receives of the order of 50-130 of those across *all* symbols (the specific amount depends on the testing set). Compared with simple canonical tasks, like balancing an inverse pendulum, which is learned from a similar (immediate) binary "yes/no" reward, this is simply much too small a sample set. For the pendulum case, reinforcement learning agents require approximately $10^7$ timesteps (number of actions in the environment) to converge. This amount of interaction with the environment is beyond the experiments we conducted: they would not reflect a realistic interaction with a human user, as they would not tolerate providing so much feedback. This illustrates the need for sample efficient methods for interactive task learning: our aim is to achieve that by exploiting the semantically rich signals in embodied conversation, and by giving the learner some control over what the learner and teacher talk about.

Each learner, except for $L_{exists}$, performs to a similar standard on $\mathcal{D}_{test}$ vs. $\mathcal{D}^*_{test}$, signifying that the developed learning models exhibit a robust

| Learner | # Queries | Existential (%) | Repeat (%) | Universal (%) |
|---|---|---|---|---|
| $L_{sem}$ | 174 | 42.1 | 15.8 | 42.1 |
| $L_{mix2}$ | 191 | 88.9 | 0 | 11.1 |
| $L_{mix1}$ | 165 | 66.7 | 9.5 | 23.8 |
| $L_{exists}$ | 238 | 88.5 | 0 | 11.5 |

**Table 7:** Summary of the number of queries and quantifiers in them (existential, repeat from the task description, or universal) used by different learners for $\mathcal{D}^*_{test}$. We do not include $L_{base}$ here because this learner does not ask any queries.

learned strategy that generalizes to new situations, even when there is concept variation.

When considering the mixture learners in the ablation experiments, $L_{mix2}$'s inferior performance to $L_{sem}$'s shows that even though they deploy the same grounder (i.e., using the logical entailments of quantifiers), the former has an inferior decision-making strategy: i.e., it hurts the learner if, during learning the action-value function $Q$, queries do not appreciate the value of information that comes from quantifiers other than *a*, such as *the one* and *every* etc. Likewise, $L_{mix1}$'s inferior performance to $L_{sem}$ shows that a better strategy cannot compensate for a grounder that does not understand the semantic difference among quantifiers.

## 5.4 Differences in the Learned Strategies

Here we explore further the different emerging behaviours of the different learners. There are summary statistics in Table 7, listing the kinds of queries the learners make during testing (on $\mathcal{D}^*_{test}$), and sample dialogues for solving the task "show me the two cyan pentagons", in the context of the same environment (but different belief states, thanks to the different grounders and learned policies of the different learners) is shown in Table 8.

Let's look first at the summary statistics in Table 7. Recall that $L_{exists}$ and $L_{mix2}$ learn to query the teacher but *without* distinguishing the expected value of queries with different quantifiers. This makes their dialogue strategy different from $L_{sem}$ and $L_{mix1}$. If acquiring the policy ignores the different consequences of different quantifiers, then our experiments show that they tend to ask queries that feature the existential quantifier in favour of alternatives: more than 88% of the queries feature an existential quantifier, which is significantly higher than their use for $L_{sem}$ and $L_{mix1}$. On the other hand,

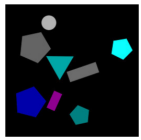
| Environment **X** | Learner | Conversation for a task |
|---|---|---|
|  | $L_{base}$ | $T$: show me the two cyan pentagons.<br>$L$: I think you are looking for this. *(points to magenta rectangle)*<br>$T$: Incorrect. |
| | $L_{exists}$ | $T$: show me the two cyan pentagons.<br>$L$: Before that, is this a cyan object *(points to blue pentagon)*<br>$T$: No.<br>$L$: Before that, show me a cyan object.<br>$T$: Here it is. *(points to a cyan pentagon)*<br>$L$: I think you are looking for this. *(points to a cyan and grey pentagons)*<br>$T$: Incorrect. |
| | $L_{mix1}$ | $T$: show me the two cyan pentagons.<br>$L$: Before that, is this a cyan object *(points to blue pentagon)*<br>$T$: No.<br>$L$: Before that, show me every cyan pentagon.<br>$T$: Here it is. *(points to all cyan pentagon)*<br>$L$: I think you are looking for this. *(points to the two cyan pentagons)*<br>$T$: Correct. |
| | $L_{mix2}$ | $T$: show me the two cyan pentagons.<br>$L$: Before that, is this a cyan object *(points to blue pentagon)*<br>$T$: No.<br>$L$: Before that, show me a cyan object.<br>$T$: Here it is. *(points to a cyan pentagon)*<br>$L$: I think you are looking for this. *(points to the two cyan pentagons)*<br>$T$: Correct. |
| | $L_{sem}$ | $T$: show me the two cyan pentagons.<br>$L$: Before that, is this a cyan object *(points to blue pentagon)*<br>$T$: No.<br>$L$: Before that, show me every cyan pentagon.<br>$T$: Here it is. *(points to all cyan pentagon)*<br>$L$: I think you are looking for this. *(points to the two cyan pentagons)*<br>$T$: Correct. |

**Table 8:** Sample conversations exhibited by different learners for solving a task of reference resolution “show me the two cyan pentagons” for the same environment.

by reasoning about the logical consequences of different quantifiers when assessing the expected value of their queries, $L_{sem}$ and $L_{mix1}$ exhibit a wider range of quantifiers in their chosen queries. This suggests that during the training process, both learners are in belief states sufficiently often that make a material difference to the expected rewards of the queries featuring different quantifiers. In particular, they must both experience a sufficient number of belief states in which the expected value of information to be gained from a query with a non-existential quantifier outweighs its higher cost (compared to the cost of a query with an existential quantifier).

The distribution of the quantifiers varies significantly among these two learners, however. That's also unsurprising: in general they will be in different belief states during training, even though they are exposed to the same images and same tasks (in the same sequence). These differences in belief, even when exposed to the same stimuli, stem from how they use evidence to build their grounders in the course of the conversations with the teacher: $L_{mix1}$ ignores the differences among the semantics of quantifiers, and so fails to obtain as much negative support for the symbols. This presumably leads to poorer quality (negative) prototypes, given $L_{mix1}$'s inferior performance on the task compared with $L_{sem}$ (see the F1-scores in Figure 5b). But when learning their policies, it also means that they are computing $Q$-values in different belief states and so learn a different mapping from the belief state to the expected reward of their options.

In the sample dialogues in Table 8, $L_{base}$ has a very cursory exchange with $T$ because it never asks questions. $L_{exists}$ and $L_{mix2}$ (i.e., the agents that don't exploit quantifier semantics when learning policies) both ask the same sequence of two queries, both using the indefinite determiner (one clarification query; one exploration query). Indeed, in spite of their different belief states at this stage—they must be different belief states, given that $L_{mix2}$ successfully executes $T$'s instruction while $L_{exists}$ doesn't—they have chosen the same actions in this exchange.

In contrast, $L_{mix1}$ and $L_{sem}$, who both exploit the logical consequences of quantifiers when learning their policies, try at first a relatively cheap way of learning more about the denotations of the symbol `cyan` (like the other learners, they must have been sufficiently uncertain about it that this query has a higher expected reward than risking executing the instruction). But on receiving the answer, the learner's belief state must still be sufficiently uncertain about how to recognise cyan objects to avoid the risk of executing the reference task. Instead, they deem it preferable to ask a query with a higher cost (featuring a universal)—in other words, the expected value of information of the query with the universal quantifier now dominates the trade-off

the learner is making when calculating which action has the highest expected reward. Even though $L_{sem}$ may at test time miscalculate the value of information it will actually gain from a response to a query—as it does in this sample dialogue because it would have been cheaper overall to simply ask the universal question in the first place—it's still the case that on average this learner performs better than the alternatives (as shown in the cumulative reward and F1-scores in Figure 5).

# 6 Conclusion

We have developed an agent that jointly learns interactive symbol grounding and a dialogue strategy for enhancing the accuracy of grounding. Our model supports incremental learning and it adapts its inferences when it discovers unforeseen possible domain states as a byproduct of interpreting the teacher's embodied utterances, which feature neologisms. Crucially, this is the first such model to explore the effects of exploiting the valid consequences of logical words like *every* and *both*: these truth conditions expand the set of training exemplars (crucially, negative exemplars) that inform grounding, and also the expected value of information for the queries the learner can ask the teacher, with a view to improving its performance on grounding and hence the reference resolution task that it faces. Our experiments demonstrate that using these logical consequences for both learning dialogue strategies and learning grounding models leads to a more data-efficient interaction, compared to learners lacking such capabilities.

There are several future directions. First, the knowledge acquisition process presented here can be complemented with additional knowledge sources, such as large language models (Wray et al., 2021) and embodied demonstrations (Argall et al., 2009), making a more comprehensive ITL system. Second, more complex tasks like rearrangements (Batra et al., 2020) could be tackled, in which the learner must not only identify referents but also manipulate them so as to reach a desired configuration. This would involve extending the types of speech acts the learner handles: for example adding the capacity to learn from the teacher's corrective feedback (Appelgren and Lascarides, 2020). Third, in order to increase the range of dialogues that a semantics-aware learner can take advantage of, we would need to incorporate more sophisticated theories of dialogue semantics into the reasoning component, in particular to handle phenomena such as co-reference resolution. Finally, our experiments deployed teachers that provide perfect information: the teacher is sincere (believes what she says) and competent (what she says

is true). In reality, even with the best of intentions, humans are not perfectly competent (and not necessarily sincere). It remains future work to investigate how teacher errors affect the learning process (but see Appelgren and Lascarides (2021) for an initial study).

## Acknowledgements

This work was supported in part by the UKRI Centre for Doctoral Training in Natural Language Processing, the UKRI (grant EP/S022481/1) and UKRI Strategic Priorities Fund to the UKRI Research Node on Trustworthy Autonomous Systems Governance and Regulation (grant EP/V026607/1, 2020-2024). Ramamoorthy is supported by a UKRI Turing AI World Leading Researcher Fellowship on AI for Person-Centred and Teachable Autonomy (grant EP/Z534833/1).

## References

Muhannad Alomari, Paul Duckworth, Majd Hawasly, David C. Hogg, and Anthony G. Cohn. 2017a. Natural language grounding and grammar induction for robotic manipulation commands. In *Proceedings of the First Workshop on Language Grounding for Robotics*, pages 35–43, Vancouver, Canada. Association for Computational Linguistics.

Muhannad Alomari, Paul Duckworth, David C. Hogg, and Anthony G. Cohn. 2017b. Natural language acquisition and grounding for embodied robotic systems. In *Thirty-First AAAI Conference on Artificial Intelligence*, pages 4349–4356. AAAI Press.

Muhannad Alomari, Fangjun Li, David C. Hogg, and Anthony G. Cohn. 2022. Online perceptual learning and natural language acquisition for autonomous robots. *Artificial Intelligence*, 303:103637.

Jacob Andreas and Dan Klein. 2016. Reasoning about pragmatics with neural listeners and speakers. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1173–1182, Austin, Texas. Association for Computational Linguistics.

Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, and Devi Parikh. 2015. VQA: Visual Question Answering. In *International Conference on Computer Vision (ICCV)*.

Mattias Appelgren and Alex Lascarides. 2020. Interactive task learning via corrective feedback. *Journal of Autonomous Agents and Multi-Agent Systems*, 34(54).

Mattias Appelgren and Alex Lascarides. 2021. Symbol grounding and task learning from imperfect corrections. In *Proceedings of the Second International Combined Workshop on Spatial Language Understanding and Grounded Communication for Robotics (SPLU-RoboNLP)*, ACL-IJNLP 2021.

Brenna D. Argall, Sonia Chernova, Manuela M. Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics Auton. Syst.*, 57(5):469–483.

Nicholas Asher and Alex Lascarides. 2013. Strategic conversation. *Semantics and Pragmatics*, 6(2):2:1–:62.

Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and S. Sabanovic. 2020. *Human-Robot Interaction: An Introduction*.

Jon Barwise and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219.

Dhruv Batra, Angel X. Chang, Sonia Chernova, Andrew J. Davison, Jia Deng, Vladlen Koltun, Sergey Levine, Jitendra Malik, Igor Mordatch, Roozbeh Mottaghi, Manolis Savva, and Hao Su. 2020. Rearrangement: A challenge for embodied AI. *CoRR*, abs/2011.01975.

Yoshua Bengio, Yann LeCun, and Geoffrey E. Hinton. 2021. Deep learning for AI. *Commun. ACM*, 64(7):58–65.

Christopher M. Bishop. 2007. *Pattern recognition and machine learning, 5th Edition*. Information science and statistics. Springer.

Norman M Bradburn and Carrie Miles. 1979. Vague quantifiers. *Public Opinion Quarterly*, 43(1):92–101.

Alan Bundy and Xue Li. 2023. Representational change is integral to reasoning. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 381.

Jean Caelen and Anne Xuereb. 2011. Dialogue and game theory. In *2011 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pages 1–10.

José Miguel Cano Santín, Simon Dobnik, and Mehdi Ghanimifard. 2020. Fast visual grounding in interaction: bringing few-shot learning with neural networks to an interactive robot. In *Proceedings of the Probability and Meaning Conference (PaM 2020)*, pages 53–61, Gothenburg. Association for Computational Linguistics.

Rich Caruana. 1997. Multitask learning. *Mach. Learn.*, 28(1):41–75.

Justine Cassell. 2001. Embodied conversational agents: Representation and intelligence in user interfaces. *AI Mag.*, 22(4):67–84.

Khyathi Raghavi Chandu, Yonatan Bisk, and Alan W. Black. 2021. Grounding 'grounding' in NLP. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 4283–4305. Association for Computational Linguistics.

Chin-Liang Chang and Richard C. T. Lee. 1973. *Symbolic logic and mechanical theorem proving*. Computer science classics. Academic Press.

Mark Chavira and Adnan Darwiche. 2008. On probabilistic inference by weighted model counting. *Artif. Intell.*, 172(6-7):772–799.

Herbert H. Clark. 1996. Using language.

Robin Cooper. 2023. *From perception to communication: a theory of types for action and meaning*. Oxford University Press.

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. 2015. Probabilistic type theory and natural language semantics. *Linguistic Issues in Language Technology*, 10.

Ann Copestake and Dan Flickinger. 2000. An open source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, Athens, Greece. European Language Resources Association (ELRA).

Ann A. Copestake, D. Flickinger, C. Pollard, and I. Sag. 1997. Minimal recursion semantics: An introduction. *Research on Language and Computation*, 3:281–332.

Gergely Csibra and György Gergely. 2009. Natural pedagogy. *Trends in Cognitive Sciences*, 13:148–153.

Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José M.F. Moura, Devi Parikh, and Dhruv Batra. 2017. Visual Dialog. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Samyak Datta, Karan Sikka, Anirban Roy, Karuna Ahuja, Devi Parikh, and Ajay Divakaran. 2019. Align2ground: Weakly supervised phrase grounding guided by image-caption alignment. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 2601–2610. IEEE.

David DeVault, Natalia Kariaeva, Anubha Kothari, Iris Oved, and Matthew Stone. 2005. An information-state approach to collaborative reference. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 1–4, Ann Arbor, Michigan. Association for Computational Linguistics.

Simon Dobnik, Robin Cooper, Adam Ek, Bill Noble, Staffan Larsson, Nikolai Ilinykh, Vladislav Maraev, and Vidya Somashekarappa. 2022. In search of meaning and its representations for computational linguistics. In *Proceedings of the 2022 CLASP Conference on (Dis)embodiment*, pages 30–44, Gothenburg, Sweden. Association for Computational Linguistics.

Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. 2023. Palm-e: An embodied multimodal language model. In *arXiv preprint arXiv:2303.03378*.

Ye Du, Zehua Fu, Qingjie Liu, and Yunhong Wang. 2021. Visual grounding with transformers. *2022 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6.

Michael C. Frank and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Daniel Fried, Justin Chiu, and Dan Klein. 2021. Reference-centric models for grounded collaborative dialogue. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2130–2147, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Drew Fudenberg and David K. Levine. 1998. The theory of learning in games.

L. T. F. Gamut. 1991. *Logic, language, and meaning. Vol.1, Introduction to logic ; L.T.F. Gamut.* University of Chicago Press, Chicago, Ill. ;.

Caelan Reed Garrett, Rohan Chitnis, Rachel Holladay, Beomjoon Kim, Tom Silver, Leslie Pack Kaelbling, and Tomas Lozano-Perez. 2021. Integrated task and motion planning. *Annual review of control, robotics, and autonomous systems*, 4.

Mario Giulianelli and Raquel Fernández. 2021. Analysing human strategies of information transmission as a function of discourse context. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 647–660, Online. Association for Computational Linguistics.

Mario Giulianelli, Arabella Sinclair, and Raquel Fernández. 2021. Is information density uniform in task-oriented dialogues? In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8271–8283, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Noah D. Goodman and Michael C. Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11):818–829.

Noah D. Goodman and Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Top. Cogn. Sci.*, 5(1):173–184.

H. P. Grice. 1975a. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics Volume 3: Speech Acts*, pages 41–58. Academic Press.

H. Paul Grice. 1975b. Logic and conversation. *Syntax and Semantics*, 3:41–58.

Janosch Haber, Tim Baumgärtner, Ece Takmaz, Lieke Gelderloos, Elia Bruni, and Raquel Fernández. 2019. The PhotoBook dataset: Building common ground through visually-grounded dialogue. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1895–1910, Florence, Italy. Association for Computational Linguistics.

Sven Ove Hansson. 2022. Logic of Belief Revision. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, Spring 2022 edition. Metaphysics Research Lab, Stanford University.

Xiaoran Hao, Yash Jhaveri, and Patrick Shafto. 2023. Common ground in cooperative communication. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.

Stevan Harnad. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346.

Robert Hawkins, Minae Kwon, Dorsa Sadigh, and Noah Goodman. 2020. Continual adaptation for efficient machine communication. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 408–419, Online. Association for Computational Linguistics.

Wilfrid Hodges. 2022. Model Theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, Spring 2022 edition. Metaphysics Research Lab, Stanford University.

Ronald A. Howard. 1966. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26.

Ronghang Hu, Marcus Rohrbach, Jacob Andreas, Trevor Darrell, and Kate Saenko. 2016a. Modeling relationships in referential expressions with compositional modular networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4418–4427.

Ronghang Hu, Huazhe Xu, Marcus Rohrbach, Jiashi Feng, Kate Saenko, and Trevor Darrell. 2016b. Natural language object retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. 2017. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 2261–2269. IEEE Computer Society.

Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2020. Compositionality decomposed: how do neural networks generalise?

Julian Jara-Ettinger, Hyowon Gweon, Laura E. Schulz, and Joshua B. Tenenbaum. 2016. The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, 20:589–604.

Doga Kisa, Guy Van den Broeck, Arthur Choi, and Adnan Darwiche. 2014. Probabilistic sentential decision diagrams. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press.

Mykel J. Kochenderfer, Tim A. Wheeler, and Kyle H. Wray. 2022. *Algorithms for Decision Makings*. The MIT Press, Boston.

Alexander Koller and Stefan Thater. 2005. The evolution of dominance constraint solvers. In *Proceedings of Workshop on Software*, pages 65–76, Ann Arbor, Michigan. Association for Computational Linguistics.

Satwik Kottur, José M. F. Moura, Devi Parikh, Dhruv Batra, and Marcus Rohrbach. 2018. Visual coreference resolution in visual dialog using neural module networks. In *The European Conference on Computer Vision (ECCV)*.

Satwik Kottur, José M. F. Moura, Devi Parikh, Dhruv Batra, and Marcus Rohrbach. 2019. CLEVR-dialog: A diagnostic dataset for multi-round reasoning in visual dialog. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 582–595, Minneapolis, Minnesota. Association for Computational Linguistics.

Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Li Fei-Fei. 2016. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *CoRR*, abs/1602.07332.

Alexander Kuhnle and Ann A. Copestake. 2017. Shapeworld - A new test methodology for multimodal language understanding. *CoRR*, abs/1704.04517.

John E. Laird, Kevin A. Gluck, John R. Anderson, Kenneth D. Forbus, Odest Chadwicke Jenkins, Christian Lebiere, Dario D. Salvucci, Matthias Scheutz, Andrea Thomaz, J. Gregory Trafton, Robert E. Wray, Shiwali

Mohan, and James R. Kirk. 2017. Interactive task learning. *IEEE Intelligent Systems*, 32(4):6–21.

Staffan Larsson. 2013. Formal semantics for perceptual classification. *Journal of Logic and Computation*, 25(2):335–369.

Staffan Larsson. 2021. The role of definitions in coordinating on perceptual meanings. In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*, Potsdam, Germany. SEMDIAL.

Staffan Larsson, Jean-Philippe Bernardy, and Robin Cooper. 2021. Semantic learning in a probabilistic type theory with records. In *Proceedings of the ESSLLI 2021 Workshop on Computing Semantics with Types, Frames and Related Structures*, pages 35–44, Utrecht, The Netherlands (online). Association for Computational Linguistics.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven C. H. Hoi. 2023. BLIP-2: bootstrapping language-image pre-training with frozen image encoders and large language models. *CoRR*, abs/2301.12597.

Xue Li, Alan Bundy, and Alan Smaill. 2018. ABC repair system for datalog-like theories. In *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2018, Volume 2: KEOD, Seville, Spain, September 18-20, 2018*, pages 333–340. SciTePress.

Vladimir Lifschitz. 2008. What is answer set programming? In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13-17, 2008*, pages 1594–1597. AAAI Press.

Sharid Loáiciga, Simon Dobnik, and David Schlangen. 2021. Reference and coreference in situated dialogue. In *Proceedings of the Second Workshop on Advances in Language and Vision Research*, pages 39–44, Online. Association for Computational Linguistics.

Robin Manhaeve, Giuseppe Marra, Thomas Demeester, Sebastijan Dumancic, Angelika Kimmig, and Luc De Raedt. 2021. Neuro-symbolic AI = neural + logical + probabilistic AI. In Pascal Hitzler and Md. Kamruzzaman Sarker, editors, *Neuro-Symbolic Artificial Intelligence: The State of the Art*, volume 342 of *Frontiers in Artificial Intelligence and Applications*, pages 173–191. IOS Press.

Jiayuan Mao, Chuang Gan, Pushmeet Kohli, Joshua B. Tenenbaum, and Jiajun Wu. 2019. The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision. In *International Conference on Learning Representations*.

Jiayuan Mao, Haoyue Shi, Jiajun Wu, Roger P. Levy, and Joshua B. Tenenbaum. 2021. Grammar-Based Grounded Lexicon Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Junhua Mao, Jonathan Huang, Alexander Toshev, Oana Camburu, Alan L. Yuille, and Kevin Murphy. 2016. Generation and comprehension of unambiguous object descriptions. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 11–20. IEEE Computer Society.

Cynthia Matuszek. 2018. Grounded language learning: Where robotics and NLP meet. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 5687–5691. ijcai.org.

Jorge A. Mendez and Eric Eaton. 2022. How to reuse and compose knowledge for a lifetime of tasks: A survey on continual learning and functional composition. *CoRR*, abs/2207.07730.

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. 2012. Foundations of machine learning. In *Adaptive computation and machine learning*.

Will Monroe, Robert X. D. Hawkins, Noah D. Goodman, and Christopher Potts. 2017. Colors in context: A pragmatic neural model for grounded language understanding. *Trans. Assoc. Comput. Linguistics*, 5:325–338.

Tim A. Wheeler Mykel J. Kochenderfer and Kyle H. Wray. 2022. *Algorithms for Decision Making*. The MIT Press, Boston Massachusetts.

Bill Noble and Nikolai Ilinykh. 2023. Describe me an auklet: Generating grounded perceptual category descriptions. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9330–9347, Singapore. Association for Computational Linguistics.

Bill Noble, Staffan Larsson, and Robin Cooper. 2022. Classification systems: Combining taxonomical and perceptual lexical meaning. In *Proceedings of the 3rd Natural Logic Meets Machine Learning Workshop (NALOMA III)*, pages 11–16, Galway, Ireland. Association for Computational Linguistics.

Charles K. Ogden, Ivor A. Richards, John Percival Postgate, Bronisław Malinowski, and F. Graham Crookshank. 1924. The meaning of meaning : a study of the influence of language upon thought and of the science of symbolism. *The Philosophical Review*, 21:212.

Aishwarya Padmakumar, Jesse Thomason, and Raymond J. Mooney. 2017. Integrated learning of dialog strategies and semantic parsing. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 547–557, Valencia, Spain. Association for Computational Linguistics.

Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.

Anna N. Rafferty, Emma Brunskill, Thomas L. Griffiths, and Patrick Shafto. 2016. Faster teaching via POMDP planning. *Cogn. Sci.*, 40(6):1290–1332.

Dan Roth. 2017. Incidental supervision: Moving beyond supervised learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 4885–4890. AAAI Press.

Rimvydas Rubavicius and Alex Lascarides. 2022. Interactive symbol grounding with complex referential expressions. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4863–4874, Seattle, United States. Association for Computational Linguistics.

Gavin Rummery and Mahesan Niranjan. 1994. On-line Q-learning using connectionist systems. Technical Report TR 166, Cambridge University Engineering Department, Cambridge, England.

Bertrand Russell. 1917. Knowledge by acquaintance and knowledge by description. In *Mysticism and Logic*, pages 152–167. London: Longmans Green.

Stuart Russell and Peter Norvig. 2020. *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson.

Patrick Shafto, Noah D. Goodman, and Thomas L. Griffiths. 2014. A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, 71:55–89.

Daniel L. Silver, Qiang Yang, and Lianghao Li. 2013. Lifelong machine learning systems: Beyond learning algorithms. In *Lifelong Machine Learning, Papers from the 2013 AAAI Spring Symposium, Palo Alto, California, USA, March 25-27, 2013*, volume SS-13-05 of *AAAI Technical Report*. AAAI.

Dan Sperber and Deirdre Wilson. 1986. Relevance: Communication and cognition.

Richard S. Sutton and Andrew G. Barto. 1998. *Reinforcement learning - an introduction*. Adaptive computation and machine learning. MIT Press.

Ece Takmaz, Nicolo' Brandizzi, Mario Giulianelli, Sandro Pezzelle, and Raquel Fernandez. 2023. Speaking the language of your listener: Audience-aware adaptation via plug-and-play theory of mind. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 4198–4217, Toronto, Canada. Association for Computational Linguistics.

Ece Takmaz, Mario Giulianelli, Sandro Pezzelle, Arabella Sinclair, and Raquel Fernández. 2020. Refer, Reuse, Reduce: Generating Subsequent References in Visual and Conversational Contexts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4350–4368, Online. Association for Computational Linguistics.

Ece Takmaz, Sandro Pezzelle, and Raquel Fernández. 2022. Less descriptive yet discriminative: Quantifying the properties of multimodal referring utterances via CLIP. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 36–42, Dublin, Ireland. Association for Computational Linguistics.

Hao Tan and Mohit Bansal. 2019. LXMERT: learning cross-modality encoder representations from transformers. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5099–5110. Association for Computational Linguistics.

Ana Tanevska, Francesco Rea, Giulio Sandini, Lola Cañamero, and Alessandra Sciutti. 2020. A socially adaptable framework for human-robot interaction. *Frontiers Robotics AI*, 7:121.

Jesse Thomason, Aishwarya Padmakumar, Jivko Sinapov, Nick Walker, Yuqian Jiang, Harel Yedidsion, Justin Hart, Peter Stone, and Raymond J. Mooney. 2020. Jointly improving parsing and perception for natural language commands through human-robot dialog. In *The Journal of Artificial Intelligence Research (JAIR)*, volume 67.

Will Thompson and Stefan Kaufmann. 2010. Signaling games with partially observable actions as a model of conversational grounding. In *Interactive Decision Theory and Game Theory, Papers from the 2010 AAAI Workshop, Atlanta, Georgia, USA, July 12, 2010*, volume WS-10-03 of *AAAI Technical Report*. AAAI.

Leslie G. Valiant. 1979. The complexity of computing the permanent. *Theor. Comput. Sci.*, 8:189–201.

Gido M. van de Ven, Tinne Tuytelaars, and Andreas S. Tolias. 2022. Three types of incremental learning. *Nat. Mac. Intell.*, 4(12):1185–1197.

Ramakrishna Vedantam, Samy Bengio, Kevin Murphy, Devi Parikh, and Gal Chechik. 2017. Context-aware captions from context-agnostic supervision. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1070–1079. IEEE Computer Society.

Ricardo Vilalta and Youssef Drissi. 2002. A perspective view and survey of meta-learning. *Artif. Intell. Rev.*, 18(2):77–95.

Pei Wang, Junqi Wang, Pushpi Paranamana, and Patrick Shafto. 2020a. A mathematical theory of cooperative communication. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Renhao Wang, Jiayuan Mao, Joy Hsu, Hang Zhao, Jiajun Wu, and Yang Gao. 2023. Programmatically grounded, compositionally generalizable robotic manipulation. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.

Sida I. Wang. 2017. *Learning adaptive language interfaces through interaction*. Ph.D. thesis, Stanford University, USA.

Sida I. Wang, Percy Liang, and Christopher D. Manning. 2016. Learning language games through interaction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2368–2378, Berlin, Germany. Association for Computational Linguistics.

Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. 2020b. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.*, 53(3):63:1–63:34.

Julia White, Jesse Mu, and Noah D. Goodman. 2020. Learning to refer informatively by amortizing pragmatic reasoning. In *Proceedings of the 42th Annual Meeting of the Cognitive Science Society - Developing a Mind: Learning in Humans, Animals, and Machines, CogSci 2020, virtual, July 29 - August 1, 2020*. cognitivesciencesociety.org.

Robert E. Wray, James R. Kirk, and John E. Laird. 2021. Language models as a knowledge source for cognitive agents. *CoRR*, abs/2109.08270.

Zhuo Yang, Yufei Han, Guoxian Yu, and Xiangliang Zhang. 2019. Prototypical networks for multi-label learning. *CoRR*, abs/1911.07203.

Linwei Ye, Mrigank Rochan, Zhi Liu, and Yang Wang. 2019. Cross-modal self-attention network for referring image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 10502–10511. Computer Vision Foundation / IEEE.

Licheng Yu, Hao Tan, Mohit Bansal, and Tamara L. Berg. 2017. A joint speaker-listener-reinforcer model for referring expressions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 3521–3529. IEEE Computer Society.

Sina Zarrieß and David Schlangen. 2019. Know what you don't know: Modeling a pragmatic speaker that refers to objects of unknown categories. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 654–659, Florence, Italy. Association for Computational Linguistics.

# A Mathematical Notation

| Symbol/Concept | Description |
|---|---|
| $u_1, u_2, \ldots$ | Entities |
| $U$ | Set of Entities (domain of discourse) |
| $\mathtt{p}, \mathtt{red}\ldots$ | Predicate symbols |
| $V$ | Vocabulary (set of known symbols) |
| $I$ | Interpretation function |
| $\mathcal{M}$ | Domain model |
| $\hat{\mathcal{M}}$ | Estimated domain model |
| $\mathtt{u}_1, \mathtt{u}_2, \ldots$ | Constants for entities |
| $a, \mathtt{red}(\mathtt{u}_1)$ | Atom constructed form symbols and constants |
| $\mathcal{H}$ | Herbrand base |
| $\mathcal{H}_{\mathcal{M}}$ | Domain model as the set of atoms |
| $\phi$ | Well-formed logical formula |
| $\vee, \wedge, \neg$ | Logical connectives |
| $\mathtt{x}$ | Variable of a logical formula |
| $\Delta$ | Domain theory (set of logical formulae) |
| `WMC` | Weighted model counting computation |
| `EVI` | Complete evidence probabilistic query |
| `MAP` | Maximum-a-posteriori probabilistic query |
| $\mathbf{w} \in [0,1]^{\mathcal{H}}$ | Weights (Bernoulli probability for each atom) |
| $r$ | Referential expression |
| $\mathcal{R}$ | Referent (set of sets of entities) |
| $\Phi(r)$ | Logical form of $r$ of a form $\langle \mathtt{Q}\ x.\phi \rangle$ |
| $\Phi(r)[\mathcal{R}]$ | Logical formulae build from $\Phi(r)$ and $\mathcal{R}$ |
| $\Phi(r)^{\mathcal{M}}$ | Referent of $\Phi(r)$ constructed using $\mathcal{M}$ |
| $\sigma(\mathcal{M}, \phi, x)$ | $\mathcal{M}$-projection ($\mathcal{M}'$ s.t. $\forall u \in U' \subseteq U\ \mathcal{M}' \models \phi[\mathtt{x}/\mathtt{u}]$) |
| $\langle \mathtt{Q} \rangle^{\mathcal{M}}$ | Referent constructor ($\langle \mathtt{Q} \rangle^{\mathcal{M}} = \{R \subseteq U \mid C_{\mathtt{Q}}(R, U)\}$) |
| $C_{\mathtt{Q}}(R, B)$ | $\mathtt{Q}$ condition between restrictor $R$ and body $B$ sets |

**Table 9:** Symbols and their descriptions used for reasoning about the domain (Section 3.1).

| Symbol/Concept | Description |
|---|---|
| $\mathbf{u} \in \mathbb{R}^d$ | $d$-dimensional feature vector for entity $u$ |
| $\mathbf{U}$ | Set of feature vectors |
| $\mathbf{y} \in \mathbb{R}^{\lvert V \rvert}$ | $\lvert V \rvert$-dimensional semantic vector |
| $\omega_{\mathcal{S}} \colon \mathbb{R}^d \mapsto [0,1]^{\lvert V \rvert}$ | Prototype network with support $\mathcal{S}$ |
| $\Omega_{\mathcal{S}} \colon \mathbb{R}^{d \times \lvert \mathbf{U} \rvert} \mapsto [0,1]^{\mathcal{H}}$ | Grounding model with support $\mathcal{S}$ |
| $\mathbb{H}$ | Bernoulli entropy |
| $\tau$ | Threshold |
| $\mathcal{S} = \{(\mathbf{u}_i, \mathbf{y}_i)\}_{i=1}^{\lvert \mathcal{S} \rvert}$ | Support of feature vector - semantic vector pairs. |
| $\mathcal{S}_{\mathtt{p}}^{+/-}$ | Positive/negative support for symbol $\mathtt{p}$ |
| $\mathbf{z}_{\mathtt{p}}^{+/-}$ | Positive/negative prototype vector for symbol $\mathtt{p}$ |
| $f \colon \mathbb{R}^d \mapsto \mathbb{R}^l$ | Encoder (Neural network feature extractor) |
| $\zeta(\mathcal{S}, \Delta)$ | Dynamic $\mathcal{S}$ update using $\Delta$ and uniform weights |

**Table 10:** Symbols and their descriptions used for interactive symbol grounding (Section 3.2).

| Symbol/Concept | Description |
|---|---|
| $L$ | Learner |
| $T$ | Teacher (domain expert) |
| $\mathbf{X} \in \mathbb{R}^{256\times256}$ | Image (ShapeWorld visual observation) |
| $q$ | query (clarification or exploration) |
| $\mathcal{Q}$ | set of queries |
| $t_r$ | Reference resolution task *"show me $r$"* |
| $\mathcal{T} = t_{r_1}, t_{r_2}, \ldots, t_{r_{\lvert\mathcal{T}\rvert}}$ | Sequence of reference resolution tasks |
| $\mathcal{C} = (\mathbf{X}, \mathcal{T})$ | Embodied conversation for task $\mathcal{T}$ in context $\mathbf{X}$ |
| $\mathcal{D} = \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_{\lvert\mathcal{D}\rvert}$ | Dataset over the sequence of embodied conversations |
| $\Phi(r)^{\hat{\mathcal{M}}}$ | Referent of $r$ estimated using $\hat{\mathcal{M}}$ |
| $\texttt{Cost}\colon \mathcal{Q} \mapsto \mathbb{R}_>$ | Query cost function |
| $c_{point} \in \mathbb{R}_>$ | Unit pointing cost |
| $c_{ref} \in \mathbb{R}_>$ | Unit reference resolution cost |
| $Ent\colon \mathcal{Q} \mapsto \mathbb{R}_>$ | Expected number of entities in $q$'s answer |
| $Sym\colon \mathcal{Q} \mapsto \mathbb{R}_>$ | Number of symbols in $q$'s referential expression |
| $s$ | Epistemic state |
| $S$ | State space |
| `Update` | State transition function |
| `ref` | action to perform reference resolution with $\hat{\mathcal{M}}$ |
| $A$ | Action space |
| $R\colon A \mapsto [-1, 1]$ | Reward function |
| $H\colon S \mapsto \mathbb{R}$ | Preference function |
| $\theta = \{\mathbf{v}, b\}$ | Preference function parameters |
| $h\colon [0,1]^{\mathcal{H}} \mapsto \mathbb{R}^m$ | $m$-statistics extractor function |
| `Softmax` | Softmax function |
| $\texttt{ANS}(q)$ | $T$'s answer to query $q$ |
| $Q \mapsto \mathbb{R}$ | Action-value function |
| $\texttt{VOI}\colon \mathcal{Q} \mapsto \mathbb{R}$ | Value of Information |
| $\Pr(\texttt{red} \mid s)$ | Probability of successful reference resolution in state $s$ |
| $\Phi(r)[\Phi(r)^{\hat{\mathcal{M}}}]$ | logical formula from the estimated referent $\Phi(r)^{\hat{\mathcal{M}}}$ |

**Table 11:** Symbols and their descriptions used in interactive reference resolution task formulation and decision-making algoritm (Section 4).