Incremental Pragmatics and Emergent Communication

Nicholas Tomlin  
Department of Computer Science  
Brown University  
Providence, RI 02912  
nicholas_tomlin@brown.edu

Ellie Pavlick  
Department of Computer Science  
Brown University  
Providence, RI 02912  
ellie_pavlick@brown.edu

Abstract

Recent work has demonstrated the ability of reinforcement learning agents to develop rich communication protocols in certain cooperative environments. Several key results have been achieved by constraining the task such that grounded communication is the only optimal strategy. We explore conditions under which groundedness may arise even when it is not strictly required by the task design. In particular, we suggest that incremental pragmatic reasoning could play a role in emergent communication, with evidence from the Task & Talk reference game.

1 Introduction

Hockett and Hockett (1960) list thirteen linguistic design features claimed to be present in all spoken languages, including semanticity, productivity, arbitrariness, and duality of patterning. We focus on two of these, semanticity and productivity, and study the extent to which they occur in the emergent communication strategies produced by reinforcement learning agents. Semanticity refers to the notion that words and phrases convey meaning, while productivity refers to the ability to create new expressions from existing linguistic items. We assume that some degree of groundedness is a prerequisite to semanticity and productivity and explore the conditions under which grounded communication may emerge.

Despite previous work which encodes explicit biases towards groundedness (Havrylov and Titov, 2017), we attempt to derive groundedness from general communicative mechanisms. We suggest that incremental pragmatic reasoning, which is a motivated mechanism of human language processing (Sedivy et al., 1999), may contribute to grounded communication in reinforcement learning agents. Our work is a direct response to Kottur et al. (2017), which claims that grounded, compositional language does not emerge naturally in multi-agent reinforcement learning environments. Kottur et al. (2017) base their claim on a series of experiments run on variations of the Task & Talk reference game, which was introduced in Das et al. (2017). Below, we analyze these results and provide an alternative explanation of why grounded communication does not emerge in their experiments. We suggest a modified, multi-task variant of Task & Talk, which is designed to serve as a test-bed for future work on emergent communication. Further, we show how incremental pragmatic reasoning may cause agents to develop grounded communication on this modified task.

2 Analysis of Task & Talk

2.1 Task Specification

Das et al. (2017) describe a cooperative reference game between two agents, Q-BOT and A-BOT, which must communicate to achieve a shared reward. A-BOT is given access to one of 64 possible
We will briefly argue that grounded strategies are no more efficient than non-grounded strategies for this task. This argument is loosely based on Crawford and Sobel (1982), which proves the optimality of groundedness as a continuous attribute of a token

\[ I(u, q) = \max_{a \in V_A} \frac{C(u, a)}{C(u)} \]  

(1)  

\[ G(P) = \sum_{u \in V_A} \frac{G(u)}{|V_A|} \]  

(2)

where \( C(u, a) \) denotes the number of co-occurrences between \( u \) and referent attribute \( a \).

We will briefly argue that grounded strategies are no more efficient than non-grounded strategies for this task. This argument is loosely based on Crawford and Sobel (1982), which proves the optimality of groundedness as a continuous attribute of a token.

\[ I(u, q) = \max_{a \in V_A} \frac{C(u, a)}{C(u)} \]  

(1)  

\[ G(P) = \sum_{u \in V_A} \frac{G(u)}{|V_A|} \]  

(2)

where \( C(u, a) \) denotes the number of co-occurrences between \( u \) and referent attribute \( a \).

We will briefly argue that grounded strategies are no more efficient than non-grounded strategies for this task. This argument is loosely based on Crawford and Sobel (1982), which proves the optimality of groundedness as a continuous attribute of a token.

\[ I(u, q) = \max_{a \in V_A} \frac{C(u, a)}{C(u)} \]  

(1)  

\[ G(P) = \sum_{u \in V_A} \frac{G(u)}{|V_A|} \]  

(2)

where \( C(u, a) \) denotes the number of co-occurrences between \( u \) and referent attribute \( a \).

We will briefly argue that grounded strategies are no more efficient than non-grounded strategies for this task. This argument is loosely based on Crawford and Sobel (1982), which proves the optimality of groundedness as a continuous attribute of a token.
of set partitioning strategies for communication in a similar but continuous domain; Kottur et al. (2017) notes that these set partitioning strategies occur in the Task & Talk reference game as well. Consider policies \( P_A : I \mapsto (a_1, a_2) \) and \( P_Q : (a_1, a_2) \mapsto \hat{w} \) which map referents to utterances and utterances to guesses, respectively, where, optimally, \( I = \hat{w} \). We may modify this policy, so that \( I \) is mapped to a different utterance \((a_1', a_2')\). We preserve the efficiency of our solution by mapping \( P_A : I' \mapsto (a_1, a_2) \) and \( P_A : I' \mapsto (a_1', a_2') \), where \( I' \) was initially mapped to \((a_1', a_2')\), and similarly for Q-Bot’s policy. Thus, the grounded and non-grounded strategies are equally optimal. Note that this argument is restricted to tabular methods, in which policies for \( I \) and \( I' \) are disentangled from all other referents and may therefore be swapped freely. However, we do observe similar partitioning behavior with the REINFORCE policy gradient algorithm (Williams, 1992); therefore, we predict that function approximation methods do not generalize across referents, permitting a similar argument.

**Reduction to 4x4 Variant.** In the full Task & Talk described in Section 2.1 A-Bot does not have access to the task \( G \). Since there are three unique tasks, however, Q-Bot reveals the task on its first turn when \(|V_Q| \geq 3\). Therefore, the experiments in Kottur et al. (2017) reduce to the 4x4 variant described above after the first dialogue turn.

### 3 Modified Task & Talk

#### 3.1 Motivations and Specification

As discussed in Section 2.2 it is possible to tweak environmental constraints so that groundedness is required for optimal communication. Here, we will pursue the emergence of grounded communication even when it is not strictly mandated by the task design. While the communication protocol from Section 2.2 used vocabulary size \(|V_A| = 4\), we predict that it should be possible to produce a one-to-one mapping between referent attributes and vocabulary tokens. To achieve these goals, we will set \(|V_A| = 8\) and focus on the simple 4x4 variant discussed in Section 2.3 above. While we have already shown that grounded communication does not emerge in this version, we propose the following modification: tasks may alternate between one and two attributes, so that \((\text{shape, color})\) and \((\text{shape})\) are both valid task specifications. Further, the length of dialogue is constrained, so that only a single token may be emitted in the one-attribute case. We do not expose Q-Bot to the task. A full list of modifications between task variants is shown in Table 1.

<table>
<thead>
<tr>
<th>Original (Kottur et al., 2017)</th>
<th>4x4 Baseline</th>
<th>4x4 Multitask</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q-Bot speaks</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Q-Bot observes ( G )</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Utility function ( U(\hat{w}) )</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Pragmatic model</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Curriculum learning</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Number of tasks</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Number of referents</td>
<td>64</td>
<td>16</td>
</tr>
<tr>
<td>Vocab size (</td>
<td>V_A</td>
<td>)</td>
</tr>
</tbody>
</table>

Table 1: Summary of Task & Talk modifications. Future work will provide deeper analysis of the role each modification plays in achieving groundedness.

#### 3.2 Incrementality and Curriculum Learning

It is immediately possible to achieve “grounded” communication in the one-attribute case, with vocabulary size \(|V_A| = 8\). Since A-Bot must be fully informative about one of eight possible attributes, a one-to-one correspondence between tokens and referents is required. We confirm this with experimental results in tabular Q-learning. However, this does not fix the set partitioning behavior in the two-attribute case. With tabular Q-learning in particular, the two cases are entirely distinct since A-Bot has different state representations. We see similar results with REINFORCE. We circumvent this issue with the following incremental pragmatic model, based on Cohn-Gordon et al. (2018), which provides a word-level alternative to the Rational Speech Acts framework (Frank and Goodman, 2012). Our model treats the stochastic policies \( \pi_Q(\hat{w} \mid s^Q_t; \theta_Q) \) and \( \pi_A(a_t \mid s^A_t; \theta_A) \)
We showed that memoryless Task & Talk forces groundedness as the only optimal strategy, but that groundedness may alternatively result from incremental pragmatic reasoning as described in Section 3.2 above. We take these results to indicate that groundedness could result from general properties of communication. We are currently working towards an incremental Q-BOT, which would allow interleaved reasoning between agents in the style of the Rational Speech Acts framework (Frank and Goodman, 2012). Based on sample calculations similar to the ones above, we predict this may lead to one-shot generalization to the two-attribute scenario. We aim to test these and similar models on larger-scale variants of the original Task & Talk model, in which both agents speak. Relatively, we hope to test these models in semi-cooperative scenarios such as negotiation, which have proven difficult for emergent communication (Cao et al., 2018).

### 3.3 Model Results

We train our model using the REINFORCE policy gradient algorithm, sharing parameters and architecture from Kottur et al. (2017), modified to the multitask 4x4 variant described above, with vocabulary size \(|V_A| = 8\). We note that the model converges to perfect accuracy. As shown by Table 2, the proposed setting leads the model to converge to perfect groundedness with high frequency, whereas alternative task designs lead to less consistently grounded strategies.

<table>
<thead>
<tr>
<th></th>
<th>4x4 Baseline</th>
<th>4x4 Multitask</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tabular Q-Learning</td>
<td>0.153</td>
<td>0.181</td>
</tr>
<tr>
<td>Tabular Q-Learning (MC)</td>
<td>0.151</td>
<td>0.182</td>
</tr>
<tr>
<td>REINFORCE</td>
<td>0.150</td>
<td>0.188</td>
</tr>
<tr>
<td>Pragmatic REINFORCE</td>
<td>0.153</td>
<td><strong>0.874</strong></td>
</tr>
</tbody>
</table>

Table 2: Mean policy groundedness scores (Eq. 2) across 100 iterations, with 10k episodes per iteration for tabular models. \(\sigma \leq 0.01\) for all models except the incrementally pragmatic REINFORCE in the multitask setting, where \(\sigma = 0.127\). A score of 1 denotes perfect one-to-one correspondence between utterances and actions and occurs in 29\% of simulations.

We will briefly demonstrate a calculation of \(P(a_t | s_t^A)\) to illustrate why the incremental pragmatic model works. Assume that curriculum learning is in place, and A-BOT and Q-BOT have achieved perfect accuracy on the one-attribute task. Let \(s_t^A\) denote an unseen two-attribute task \((\text{red, square})\). In this case, \(\pi_A(a_t | s_t^A; \theta_A)\) is a uniform distribution, so we consider the expected continuation term. Note that \(\pi_Q(\hat{w} | s_t^Q; \theta_Q)\) is also a uniform distribution for all two-attribute \(s_t^Q\). However, when \(s_t^Q\) contains the single attribute \(a_t\), then \(\pi_Q\) predicts a single attribute \(\hat{w}\) such that \(U(\hat{w}) = 1\). In this way, the model selects either the token corresponding to \((\text{red})\) or the token corresponding to \((\text{square})\), as desired. The model calculation for A-BOT’s second token is symmetric to the one described here.

Note that this model does not require strict grounding of communication to achieve perfect accuracy, but prefers it due to the curriculum learning method. For example, if we have one-attribute mappings \(\text{red} \rightarrow 1\) and \(\text{square} \rightarrow 5\), we could force a non-grounded mapping, e.g., \((\text{red, square}) \rightarrow (6, 7)\), by manually manipulating the \(\pi_A\) term. We expect the system to continue to exhibit grounded behavior in spite of such adversarial mappings. Exploring this is part of ongoing work.

### 4 Conclusion and Ongoing Work

We showed that memoryless Task & Talk forces groundedness as the only optimal strategy, but that groundedness may alternatively result from incremental pragmatic reasoning as described in Section 3.2 above. We take these results to indicate that groundedness could result from general properties of communication.
References


