Incremental Pragmatics and Emergent Communication

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

Abstract

Recent work has demonstrated the ability of reinforcement learning agents to develop rich communication protocols in certain cooperative environments. However, it has remained unclear how and to what extent these communication strategies generalize to unseen referents. We argue that the constraints imposed on these multi-agent environments crucially determine the structure and properties of emergent communication. In particular, we show that incremental pragmatic reasoning may lead to productivity, with evidence from the *Task & Talk* reference game (Das et al., 2017).

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. Intuitively, we might expect that reinforcement learning agents without such an explicit bias would fail to produce grounded communication. However, we suggest that incremental pragmatic reasoning, which is a well-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.

On the basis of this analysis, 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, leading to both semanticity and productivity as desired.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

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 referents, which is described in terms of three attributes (shape, color, and style). Q-BOT's task is to determine two of these attributes (e.g., color and style) via communication with A-BOT.

Agents communicate over two rounds, with Q-BOT going first. During a round, each agent may emit a single token from its vocabulary, which has a predetermined size. After A-BOT's second utterance, Q-BOT guesses two attributes from the full set of 12 possible attributes (4 color, 4 shape, 4 style).

We follow Kottur et al. (2017)'s notation. Let G refer to Q-BOT's task, and I refer to the object instance. Let $s_Q^t = [G, q_1, a_1, \ldots, q_{t-1}, a_{t-1}]$ denote the state of Q-BOT at time t, and let $s_A^t = [I, q_1, a_1, \ldots, q_{t-1}, a_{t-1}, q_t]$ denote the state of A-BOT, where q_i and a_i are tokens from vocabularies V_Q and V_A, respectively. Let $\hat{w}$ refer to Q-BOT's guess at the end of dialogue.

Consider the example dialogues in Figure 1 below:

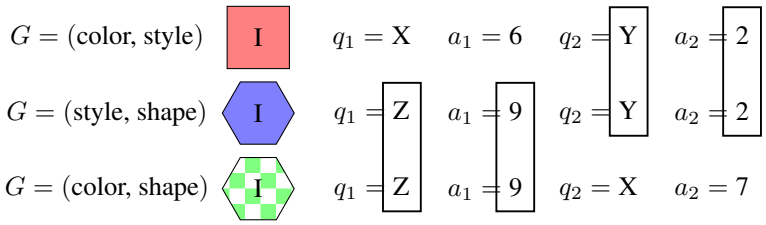

Figure 1: Example dialogues from the *Task & Talk* reference game. Note that since the first two referents share the same style (solid) the dialogues use consistent tokens $q_2 = \text{Y}$ and $a_2 = 2$ to refer to this attribute. Q-BOT's utterance "Y" may be interpreted as asking about the style of G, while A-BOT's response "2" may be interpreted as the answer "solid." Note that a similar pattern exists between the last two referents, but with a shared shape (hexagon) instead of style. This idealized dialogue is an example of grounded communication, since there is a one-to-one correspondence between referent attributes and dialogue tokens.

2.2 Simplified 4x4 Variant

Consider a simplified version of this task, in which A-BOT has access to the task specification G. In other words, A-BOT has full knowledge of Q-BOT's state representation. Because of this, Q-BOT's utterances may be functionally ignored. Therefore, the task reduces to A-BOT describing two attributes with two tokens and Q-BOT guessing these attributes correctly. In light of this, we model the 4x4 configuration with just 16 referents and no explicit task G. Rather, A-BOT is tasked with being fully informative about the referent object I. As usual, A-BOT is allowed two dialogue turns; Q-BOT, however, does not communicate and is restricted to a guessing role (since Q-BOT has no privileged information). Therefore, when $|V_A| \geq 4$, it is clear that a perfect communication strategy exists. Since A-BOT may choose from $|V_A| \cdot |V_A| \geq 16$ possible utterances, it is possible to produce a unique and fully informative mapping between utterances and referents.

We consider a communication strategy to be *grounded* if referent attributes consistently co-occur with certain vocabulary tokens (e.g., red shapes are consistently described with token 3). We measure the groundedness of a token u, and the groundedness score of a policy P, as follows:

$$G(u) = \frac{\max_a(C(u,a))}{C(u)} \qquad (1) \qquad\qquad G(P) = \sum_{u \in V_A} \frac{G(u)}{|V_A|} \qquad (2)$$

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 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 and P_Q which map referents to utterances and utterances to guesses, respectively. Thus, we can write $P_A : I \mapsto (a_1, a_2)$ and $P_Q : (a_1, a_2) \mapsto \hat{w}$ where optimally $I = \hat{w}$. We may now 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. This swapping method allows us to adjust the groundedness of a communication policy without affecting its efficiency, so that grounded and non-grounded strategies may be 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, allowing us to make a similar swapping argument to the one above.

2.3 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.

2.4 State Modifications and Memoryless Variants

Das et al. (2017) and Kottur et al. (2017) both achieve grounded communication on this task by modifying the state representation. Specifically, Das et al. (2017) removes Q-BOT's task representation G during the guessing phase, and Kottur et al. (2017) restricts A-BOT's state $s_A^t = [I, q_t]$ to only include the most recent utterance from Q-BOT. We will focus on the latter modification, although the motivations behind these two approaches are similar.

By removing A-BOT's memory, Kottur et al. (2017) ensures that the grounded communication strategy is the only optimal one. Briefly, this is because Q-BOT may convey a single trit of information per turn; if Q-BOT is fully informative about G, it cannot also convey turn information. Because of this, Q-BOT's first utterance must narrow the set of possible two-attribute guesses from $3 \cdot 4^2 = 48$ down to $2 \cdot 4^2 = 32$ (e.g., all those which distinguish color). Then, A-BOT's response must further narrow possible guesses by signaling the single attribute (e.g., the color value) which is distinguished in every possible guess, and so on.

3 Modified *Task & Talk*

3.1 Motivations and Specification

As discussed in Section 2.4, it is possible to tweak environmental constraints so that groundedness is required for optimal communication. Below, 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 2.4 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 (which Kottur et al. (2017) refer to as an attribute-value vocabulary). To achieve these goals, we will set $|V_A| = 8$ and focus on the simple 4x4 variant discussed in Section 2.2 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 *(shape, color)* and *(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 below.

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.

	Original (Kottur et al., 2017)	4x4 Baseline	4x4 Multitask
Q-BOT speaks	✓	✗	✗
Q-BOT observes G	✓	✓	✗
Utility function $U(\hat{w})$	✗	✗	✓
Pragmatic model	✗	✗	✓
Curriculum learning	✗	✗	✓
Number of tasks	3	1	3
Number of referents	64	16	16
Vocab size $\lvert V_A \rvert$	4	4	8

Table 1: Summary of *Task & Talk* modifications

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, using the model architecture described in Kottur et al. (2017). 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 $\pi_Q(\hat{w} \mid s_t^Q; \theta_Q)$ and $\pi_A(a_t \mid s_t^A; \theta_A)$ from the REINFORCE model as base agents in its pragmatic reasoning calculation:

$$P(a_t \mid s_t^A) \propto \pi_A(a_t \mid s_t^A; \theta_A) \cdot \mathbb{E}_{s_0^Q}\left[\pi_Q(\hat{w} \mid s_0^Q; \theta_Q) \cdot U(\hat{w})\right] \tag{3}$$

where $a_t \in s_0^Q$ denotes an estimation over all Q-BOT states containing a_t, and $U(\hat{w})$ denotes the utility of the guess. Crucially, we define $U(\hat{w}) = 1$ if true *or* partially true, and $U(\hat{w}) = 0$ otherwise. That is, if $\hat{w} =$ *(red)* and the target is *(red, square)*, $U(\hat{w}) = 1$.

To achieve groundedness, we use a curriculum learning method in which agents are presented with easier tasks first; in particular, one-attribute tasks are presented before two-attribute tasks, allowing Q-BOT to develop a policy for the one-attribute case.

Note that, for the sake of implementation, we use Q-BOT's policy π_Q in the model $P(a_t \mid s_t^A)$ from which A-BOT samples. Since A-BOT has full access to Q-BOT's state, this policy could alternatively be modeled via simple observation of Q-BOT's behavior, achieving the exact same results.

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, and a one-to-one correspondence between A-BOT's tokens and referent attributes emerges in the curriculum learning setting. As shown by Table 2, this method does not always converge to perfect groundedness, but does so with high relative frequency.

We will briefly demonstrate a calculation of $P(a_t \mid s_t^A)$ 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 *(red, square)*. In this case, $\pi_A(a_t \mid s_t^A; \theta_A)$ is a uniform distribution, so we consider the expected continuation term. Note that $\pi_Q(\hat{w} \mid s_t^Q; \theta_Q)$ is also a uniform distribution for all two-attribute s_t^Q. However, when s_t^Q contains the single attribute a_t, then π_Q predicts a single attribute $\hat{w}$ such that $U(\hat{w}) = 1$. In this way, the model selects either the token corresponding to *(red)* or the token corresponding to *(square)*, as desired. The model calculation for A-BOT's second token is symmetric to the one described here.

It is worth noting that this model does not *require* strict grounding of communication to achieve perfect accuracy, but prefers it due to the curriculum learning method outlined above. For example, if we have one-attribute mappings red $\rightarrow 1$ and square $\rightarrow 5$, we could force a non-grounded mapping, e.g., (red, square) $\rightarrow (6, 7)$, by manually manipulating the π_A term. While not extensively tested, we expect the rest of the system to continue to exhibit grounded and productive behavior in spite of such adversarial mappings. We consider this robustness to be an important aspect of productivity.

	4x4 Baseline	4x4 Multitask
Tabular Q-Learning	0.153	0.181
Tabular Q-Learning (MC)	0.151	0.182
REINFORCE	0.150	0.188
Pragmatic REINFORCE	0.153	**0.874**

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

4 Conclusion and Ongoing Work

Through extensive study of the *Task & Talk* reference game, we have shown how various communicative constraints and mechanisms may lead to drastically different systems of emergent communication. In particular, 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, and semanticity and productivity by extension, should 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. Relatedly, we hope to test these models in semi-cooperative scenarios such as negotiation, which have proven difficult for emergent communication (Cao et al., 2018).

References

Cao, K., Lazaridou, A., Lanctot, M., Leibo, J. Z., Tuyls, K., and Clark, S. (2018). Emergent communication through negotiation. *arXiv preprint arXiv:1804.03980*.

Cohn-Gordon, R., Goodman, N. D., and Potts, C. (2018). An incremental iterated response model of pragmatics. *arXiv preprint arXiv:1810.00367*.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451.

Das, A., Kottur, S., Moura, J. M., Lee, S., and Batra, D. (2017). Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2970–2979. IEEE.

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

Havrylov, S. and Titov, I. (2017). Emergence of Language with Multi-Agent Games: Learning to Communicate with Sequences of Symbols. In *Advances in Neural Information Processing Systems*, pages 2149–2159.

Hockett, C. F. and Hockett, C. D. (1960). The Origin of Speech. *Scientific American*, 203(3):88–97.

Kottur, S., Moura, J., Lee, S., and Batra, D. (2017). Natural Language Does Not Emerge 'Naturally' in Multi-Agent Dialog. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2962–2967.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., and Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2):109–147.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.