Emergent Compositionality in Signaling Games

Nicholas Tomlin

A thesis submitted in partial fulfillment of the requirements
for the degree of A.B. in Linguistics with Honors

Brown University
Providence, RI
May 2019

Abstract

Recent work in natural language understanding (Mordatch and Abbeel, 2017; Hermann et al., 2017) has used deep learning to generate artificial languages in simulated environments, termed *emergent communication.* We leverage these computational environments as a testing ground for theories concerning the emergence of linguistic compositionality and compare our results with human behavior on related iterated signaling games. In particular, we provide experimental evidence suggesting that incremental pragmatic reasoning may lead to compositional referring behavior in both computational agents and in humans.

Acknowledgments

First and foremost, I would like to thank my advisor Ellie Pavlick for her insightful comments and guidance throughout the duration of this project. I would also like to thank the members of the Brown NLU Group and Brown CLPS Department for their comments and help along the way. Conversations with Kenny Smith, Brenden Lake, Chris Potts, Dan Klein, and many others also helped shape the direction and framing of this work. Finally, many thanks to my experimental subjects, without whom empirical results would not be possible.

Previous Publication of This Work

Much of the work contained in this thesis has been published or will be published in other venues. In particular, Tomlin and Pavlick (2018) details our initial computational results in emergent communication, and the forthcoming Tomlin and Pavlick (2019) compares these results to experiments on human subjects. Compared to these articles, this thesis contains additional introductory material and further details concerning experimental results and statistical ablations.

Contents

Chapter 1

Introduction

Hockett and Hockett (1960) list thirteen linguistic design features claimed to occur in all spoken languages, including semanticity, productivity, arbitrariness, and duality of patterning. Following a tradition of emergentism (O'Grady, 2008), we seek to locate underlying cognitive mechanisms which may result in such design features. In particular, we focus on the emergence of *linguistic compositionality* and its underlying mechanisms.

Gibson et al. (2019) frames the emergence of compositionality as a fundamental question at the intersection of learnability and efficiency in language. With the relatively recent resurgence of deep learning methods in computational linguistics, many have argued that machine learning tools may be key to understanding the role of learnability in language (Pater, 2019; Potts, 2019; Linzen, 2019). In light of this, we consider *emergent communication*, which involves the simulation of artificial languages by computational agents, to be a particularly fruitful testing ground for this line of research. While emergent communication is certainly not a new field (Wagner et al., 2003), the recent influence of deep learning has led to increasingly complex artificial language generation (Mordatch and Abbeel, 2017; Hermann et al., 2017), enabling new methods of linguistic analysis.

Beginning with Tomlin and Pavlick (2018), we analyze emergent communication protocols in an iterated reference game setting. Put succinctly, this task involves two participants (a *sender* and a *receiver*) who repeatedly refer to some target objects in the context of relevant distractors until converging to a mutually beneficial communication strategy. Similar reference games have historically been used to model certain linguistic and pragmatic behaviors in humans (Krauss and Weinheimer, 1964; Hawkins et al., 2017). To ground our computational results in human behavior, we run a near-identical iterated signaling task on human subjects and compare their strategies against the various emergent communication protocols.

Below, we present background information on linguistic compositionality and suggest several potential mechanisms which might have led to its development. We also present background information related to these proposed mechanisms, including a primer on noisy-channel models and computational pragmatics. We then summarize past work in the field of emergent communication, as well as its human counterpart *iterated learning*. Finally, we briefly consider relevant prior work on game-theoretic pragmatics and strategic information transmission.

1.1 Background

1.1.1 Compositionality

Linguistic compositionality, or the ability to productively generate and understand the meaning of new utterances from their component parts, is regarded as a universal property of languages (Frege, 1892). Despite this, it is possible to imagine a communication system which is not compositional (De Beule and Bergen, 2006; Gibson et al., 2019). We focus on a subset of compositionality, herein referred to as *compositional referring behavior*. Below, we will define this concept and briefly explain how it differs from classical definitions of compositionality, which are presented in Section 1.1.1.1. We will then in Section 1.1.1.3 present general cognitive mechanisms which may lead to compositional referring behavior.

1.1.1.1 Classical Definitions

Partee (1984) defines compositionality in terms of syntactic composition. In particular, *the principle of compositionality* states that the meaning of a syntactic object is determined by the semantic composition of its children. This is commonly notated as $\sigma(a \oplus b) = \sigma(a) \oplus \sigma(b)$ (Goldberg, 1995), or in terms of function application as in Figure 1.1 (Potts, 2019).

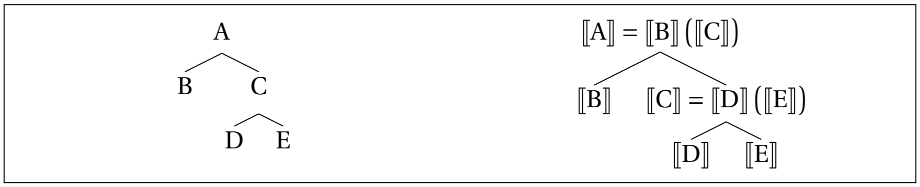


Figure 1.1: Compositionality via Function Application

Following prior work on the emergence of compositionality by Nowak and Krakauer (1999) and Brighton (2002), we do not consider the full principle of compositionality. This is largely a result of the simplicity of the iterated reference game, which does not exhibit a clear syntax-semantics distinction.

1.1.1.2 Compositionality & Groundedness

In this work, we consider a reference game scenario in which individuals refer to objects with particular attributes. In this setting, we consider compositional referring behavior to be a strategy in which individuals describe objects in terms of their component attributes. This is consistent with definitions of compositionality in related work (Kirby et al., 2014, 2015). Note that such a communication strategy is intuitively more likely to generalize to unseen objects with familiar attributes than one which is not compositional.

Groundedness is a similar concept, referring to the extent to which individual words or symbols correspond to elements of the real world (Barsalou, 2008). In this scenario, we consider a communication strategy to be grounded if referent attributes consistently co-occur with certain vocabulary tokens. Given a corpus of objects and their descriptions, we can measure groundedness as a continuous attribute of a single token u or of the entire communication policy P, as follows:

$$G(u) = \frac{\max_a(C(u,a))}{C(u)} \qquad (1.1) \qquad\qquad G(P) = \sum_{u \in V} \frac{G(u)}{|V|} \qquad (1.2)$$

where $C(u,a)$ denotes the number of co-occurrences between u and referent attribute a in the corpus, $C(u)$ denotes the total count of occurrences of u across all referent attributes, and $|V|$ denotes the size of the vocabulary. Therefore, $G(u)$ approximates the probability of correctly guessing a referent attribute corresponding to token u, and $G(P)$ averages this probability over all tokens in the corpus. We will hereafter refer to $G(P)$ as the *groundedness score* of a particular communication policy and will use this as an evaluation metric of compositional referring behavior in future sections.

1.1.1.3 Four Mechanisms Leading to Compositionality

Synthesizing prior work on the emergence of compositionality, we identify four major mechanisms leading to compositionality, which are summarized in Box 1. The first of these, *iterated transmission effects*, refers to the ability of compositional behavior to emerge over time as communication protocols are transferred from one dyad of speakers to the next; Kirby et al. (2014) demonstrates this effect experimentally in an iterated learning paradigm.

Second, *compression effects* refers to a functional pressure to condense communication systems, e.g., due to constraints on memory. After a certain threshold of use frequency defined in Yang (2016), compositional strategies become significantly more efficient than their non-compositional counterparts. We briefly consider how compression effects may be realized in computational models, and we show how they predict human behavior on the iterated reference game task.

Third, the noisy-channel model (Gibson et al., 2013) and related accounts of noise effects have been proposed as potential causes of linguistic productivity (Nowak and

Box 1: Mechanisms of Compositionality

We identify four potential general cognitive mechanisms which may cause or otherwise be prerequisite to compositional referring behavior:

1) Iterated transmission effects (Smith et al., 2003; Kirby et al., 2014)
2) Compression effects (Kirby et al., 2015; Yang, 2016)
3) Noisy-channel model (Nowak and Krakauer, 1999; Futrell, 2017)
4) Pragmatic reasoning (Tomlin and Pavlick, 2018)

These mechanisms are not necessarily incompatible with one another, nor are they necessarily sufficient for the full principle of compositionality as described in Section 1.1.1.1.

Krakauer, 1999) and syntactic structure (Futrell, 2017). This model states that if only part of a message is lost, then the whole can potentially be retrieved via rational inference over potential sources of noise. Indeed, we might expect this to make compositionality optimal: if the meaning of an utterance is compositionally distributed across its subparts, some of which are noisily omitted or altered, then some portion of the utterance meaning should still be retained. We will provide more information about these models in Section 1.1.2.

Finally, Tomlin and Pavlick (2018) proposes incremental pragmatic reasoning as a potential mechanism leading to linguistic compositionality. In the following sections, we will present motivations and experimental data supporting this mechanism. To aid this analysis, we will rigorously define incremental pragmatics in terms of the model proposed in Cohn-Gordon et al. (2018b).

1.1.2 Noisy Channel Models

Following Shannon (1948), various attempts have been made to integrate information theoretic concepts with language processing (Gibson et al., 2019). Notable among these is the noisy-channel model of sentence processing (Gibson et al., 2013), under which interlocutors rationally compute the meanings of noisy utterances based on their semantic expectations.

Futrell (2017) demonstrates how the related concept of *noisy-context surprisal* may act as a mechanism leading to dependency locality in natural languages. Futrell (2017) additionally shows how this noisy-channel model may be integrated into planning algorithms via *incremental sequence samplers* and briefly argues that such algorithms may lead to naturalistic syntactic phenomena in emergent communication systems.

1.1.3 Computational Pragmatics

1.1.3.1 Gricean Pragmatics

Pragmatic language use has traditionally been modeled in terms of Grice (1975)'s cooperative principle. The cooperative principle is not meant to be prescriptive of speakers, but rather represents a general guideline from which interlocutors may derive implicatures. It is typically subdivided into four maxims, listed below:

1. Maxim of Quality: "Try to make your contribution one that is true."
   (a) "Do not say what you believe to be false."
   (b) "Do not say that for which you lack adequate evidence."
2. Maxim of Quantity
   (a) "Make your contribution as informative as is required."
   (b) "Do not make your contribution more informative than is required."
3. Maxim of Relation: "Be relevant."
4. Maxim of Manner: "Be perspicuous."
   (a) "Avoid obscurity of expression."
   (b) "Avoid ambiguity."
   (c) "Be brief (avoid unnecessary prolixity)."
   (d) "Be orderly." (Grice, 1975)

Speakers often flout these maxims in order to convey additional meaning in the form of pragmatic implicatures. For example, if asked "When exactly is the final paper due?" a speaker might reply "Sometime next week." This fails to fulfill the maxim of quantity, suggesting that the responder has some reason for being uninformative in their answer. The listener might therefore infer that the speaker does not know the exact due date, thereby achieving the desired pragmatic effect. Examples such as these form the basis for classical pragmatic theory, but the algorithmic mechanism by which interlocutors derive these implicatures is somewhat unclear. We will therefore consider an operationalization of Grice (1975)'s maxims via computational pragmatics in the RSA model.

1.1.3.2 Rational Speech Acts (RSA) Model

The Rational Speech Acts model (Frank and Goodman, 2012) provides a computational alternative to Grice (1975)'s maxims by modeling a recursive reasoning process between speakers and listeners about utterances and their intentions. This recursive process is defined in terms of Bayesian probabilistic reasoning about a layered "stack" of speaker and listener agents, which are modeled as probability distributions $(u \mid w)$ and $(w \mid u)$, respectively, where w denotes a world state and u denotes an utterance.

In particular, we will define base speaker and listener agents S_0 and L_0 which interpret utterances literally. That is, upon hearing an utterance u, the base listener L_0 probabilistically interprets the world state to be any w such that $[[u]](w) = 1$ is consistent with u, multiplied by a prior $P(w)$ over possible world states. Meanwhile, the base speaker S_0 chooses any utterance which is compatible with the current world state, but with a preference for shorter utterances, as enforced by a cost function $C(u)$. On top of these base agents, we define pragmatic agents S_n and L_n which reason about the base agents by normalizing over the set of potential alternative utterances and world states. For example, the pragmatic listener L_1, upon hearing an utterance u, considers the set of all possible world states w and calculates the probability that the base speaker S_0 would have produced u given that world state. We may develop corresponding equations as follows:

$$P_{L_0}(w \mid u) \propto \llbracket \mathrm{u} \rrbracket(w) \cdot P(w) \tag{1.3}$$

$$P_{S_1}(u \mid w) \propto e^{\lambda\left(\log P_{L_0}(w \mid u) - C(u)\right)} \tag{1.4}$$

$$P_{L_2}(w \mid u) \propto P_{S_1}(u \mid w) \cdot P(w) \tag{1.5}$$

$$\vdots$$

$$P_{S_0}(u \mid w) \propto e^{\lambda_1\left(\log(\llbracket \mathrm{u} \rrbracket(w)) - C(u)\right)} \tag{1.6}$$

$$P_{L_1}(w \mid u) \propto P_{S_0}(u \mid w) \cdot P(w) \tag{1.7}$$

$$P_{S_2}(u \mid w) \propto e^{\lambda_2\left(\log P_{L_1}(w \mid u) - C(u)\right)} \tag{1.8}$$

$$\vdots$$

Worked examples illustrating this pragmatic reasoning process may be found in Monroe and Potts (2015) as well as Goodman and Frank (2016). Note that λ_i denote temperature parameters which may tweak the optimality of pragmatic reasoning. Further note that although we might consider a theoretically infinite stack of speakers and listeners which eventually converge (Levy, 2018), in practice only a few layers are needed to obtain pragmatic language behavior.

In recent years, the RSA model has been widely successful, and theoretical results have shown how it may account for scalar implicature (Goodman and Frank, 2016), M-implicature (Bergen et al., 2016), metaphor (Kao et al., 2014a), and hyperbole (Kao et al., 2014b), among other pragmatic phenomena. Despite this theoretical success, its application has been largely restricted to reference game settings where the set of possible alternative utterances and world states is quite small. We consider the incremental variant, where normalization occurs over alternative words rather than alternative utterances, to be a more scalable and realistic implementation of this model and present its details in the following section.

1.1.3.3 Incremental Pragmatics

Experimental work has suggested that humans perform pragmatic reasoning incrementally during processing rather than solely at the sentence-level (Sedivy et al., 1999). To operationalize this notion, Cohn-Gordon et al. (2018b) proposes an incremental variant of the standard RSA model. While this model offers slightly different predictions from classic RSA (e.g., in terms of over-informativity), it places a bound on the theoretical set of possible alternatives. That is, instead of considering the unbounded set of possible alternative sentences, we need only consider the bounded set of possible alternative words corresponding to each word in the sentence. Cohn-Gordon et al. (2018b) formalizes this process in the following manner:

$$P_{L_0}(w \mid c, \text{word}) \propto [\![\text{c} + \text{word}]\!](w) \cdot P(w) \tag{1.9}$$

where c denotes a partial utterance and $[\![\text{c} + \text{word}]\!](w)$ denotes the fraction of potential utterance continuations of $(c + \text{word})$ which are consistent with the world state w. As in Cohn-Gordon et al. (2018a), this fraction is typically approximated with a probabilistic model. Other speaker and listener agents are defined similarly to the classical RSA model.

1.2 Prior Work

1.2.1 Emergent Communication

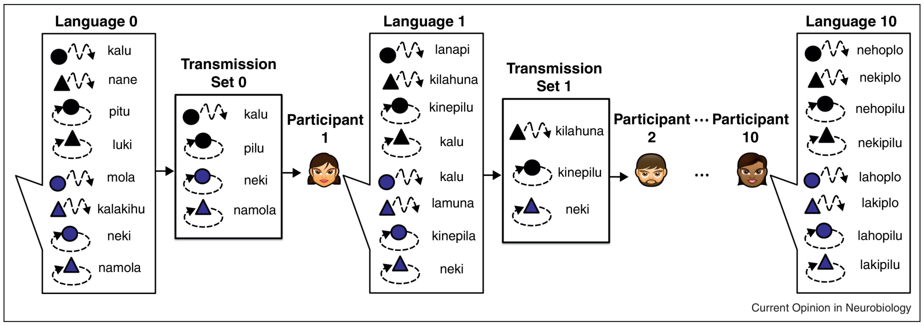
Emergent communication originated at the intersection of research in multi-agent systems and language origins and evolution (Wagner et al., 2003) and has more recently blossomed as a subfield of deep reinforcement learning (Mnih et al., 2015; Li, 2017). At the highest level, this field involves the automated development of communication protocols between autonomous, collaborative agents, although non-cooperative variants also exist (Cao et al., 2018).

Emergent communication systems may be divided into those which produce discrete (Kottur et al., 2017; Das et al., 2017; Mordatch and Abbeel, 2017) versus continuous (Foerster et al., 2016; Sukhbaatar et al., 2016) communication protocols. In the continuous setting, agents communicate by transferring vectors of real numbers; while this framework aligns well with current gradient descent methods in deep learning, the nascent communication is less interpretable and shows less parallels to human language. Discrete settings, on the other hand, typically involve a fixed-size vocabulary. While vocabulary tokens are initially “meaningless,” agents may learn to ground them to specific concepts and communicative goals over time. We will focus on the discrete domain in the following sections.

Most relevant here is the work of Kottur et al. (2017), which claims that natural language-like protocols do not occur in emergent communication systems. This argument is predicated on a series of experiments with the *Task & Talk* reference game, suggesting that emergent communication systems do not produce compositional policies in the sense of Section 1.1.1.2. We will argue that compositional policies in fact *do* emerge in this setting but require specific inductive biases to be encoded into the reinforcement learning models.

1.2.2 Iterated Learning

Prior research in iterated learning has directly addressed questions about the origins of compositionality. Most notably, Kirby et al. (2014) demonstrates how iterated transmission effects in a *telephone*-like game might lead to compositional referring behavior. This experiment, visualized in Figure 1.2, involves a set of eight referents which participants described to one another over the course of ten dyads. While the initial referring expressions are purposefully non-compositional, the final communication policies show strong signs of compositionality, with specific morphemes corresponding to particular referent attributes.



Current Opinion in Neurobiology

Figure 1.2: Emergence of compositionality in telephone game with restricted transmission over multiple dyads. Reproduced from Kirby et al. (2014).

Relatedly, Kirby et al. (2015) uses an iterated learning paradigm to demonstrate a trade-off between the expressivity and compressibility of communication protocols, where compressibility is measured in terms of the coding length of a minimally redundant grammar.

1.2.3 Game Theoretic Approaches

Many similar questions have been addressed in the subfield of game theory known as *strategic information transmission* (Crawford and Sobel, 1982). This type of analysis is typically restricted to very simple communicative domains, and it is difficult to generalize game theoretic results to more complex settings. Nevertheless, Crawford and Sobel (1982) proves that in a specific continuous setting, all communication strategies which partition the real-valued signaling space into subspaces of equal size are equally optimal. This result is a precursor to discrete *partitioning strategies* which will be discussed in the context of Kottur et al. (2017)'s *Task & Talk* paradigm.

Chapter 2

Computational Experiments

We run a variety of computational experiments using the emergent communication paradigm. These experiments primarily involve pairs of reinforcement learning agents which eventually converge to mutually beneficial communication strategies in the *Task & Talk* task, which is described in Sections 2.1.1 and 2.1.3. These computational agents are parametrized by standard reinforcement learning algorithms, which are summarized and briefly explained in Section 2.1.2.

2.1 Background & Methods

2.1.1 Task Specification

We focus on the *Task & Talk* reference game task, which was originally defined in Das et al. (2017) and further studied in Kottur et al. (2017). We will first consider the task as presented in Das et al. (2017) and then propose a variety of well-motivated modifications in Section 2.1.3.

Das et al. (2017) describes 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 a fixed-size vocabulary. 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). Let G refer to Q-BOT's task, and I refer to the object instance. Let $s_Q^t = [G, q_1, a_1, \dots, q_{t-1}, a_{t-1}]$ denote the state of Q-BOT at time t, and let $s_A^t = [I, q_1, a_1, \dots, q_{t-1}, a_{t-1}, q_t]$ 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.

This signaling task is repeated indefinitely, until the agents converge to an optimal communication strategy. Note that *optimal* in this sense does not refer to the groundedness of the communication policy but rather its ability to correctly and consistently signal referents in the training set. Once a policy has converged to a stable state, we may extract sample dialogues and calculate a groundedness score. We show sample dialogues in Figure 2.1 below.

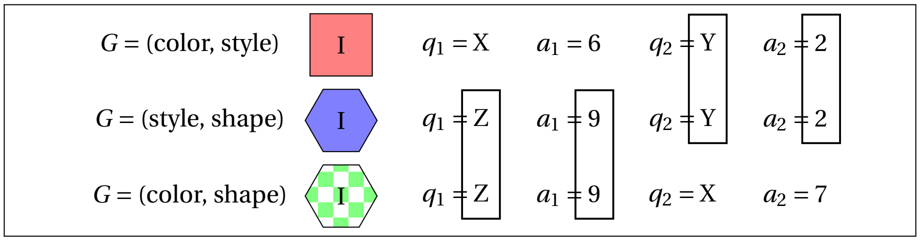


Figure 2.1: Example dialogues from the *Task & Talk* reference game. Note that since the first two referents share the same style (solid) the dialogues use consistent tokens q_2 = Y and a_2 = 2 to refer to this attribute. "Y" may be interpreted as asking about the style of G, while "2" may be interpreted as the answer "solid." This idealized dialogue is an example of grounded communication, since there is a one-to-one correspondence between referent attributes and dialogue tokens.

2.1.2 Reinforcement Learning

Before moving on, we will briefly review key algorithms in reinforcement learning. Roughly, these algorithms all perform the same task of mapping world states to actions, and they require past (state, action, reward) information in order to maximize the potential of future reward. Recall that in this setting, world states s_Q^t and s_A^t denote task and referent information, along with existing dialogue history. Actions q_i and a_i denote single-token utterances.

However, reinforcement learning algorithms differ in how they produce these mappings ($s_Q^t \to q_i$) and ($s_A^t \to a_i$). Tabular methods, discussed in Sections 2.1.2.1 and 2.1.2.2, store complete reward histories in large tables; typically, actions are chosen in order to maximize the likelihood of receiving a reward. On the other hand, policy gradient methods such as REINFORCE (Section 2.1.2.3) generate probability distributions over actions given states. These probability distributions are often approximated with deep neural networks, allowing the algorithm to generate actions for previously unseen world states.

2.1.2.1 Tabular Q-Learning

Tabular Q-learning methods seek to optimize a Q-function which maps state-action pairs to expected reward values as follows:

$$Q : S \times A \to \mathbb{R} \tag{2.1}$$

where S and A denote the sets of world states and potential actions, respectively. This function, often represented as a Q-table, may be defined according to the following update step:

$$Q(s_t, a_t) \leftarrow (1-\alpha) \cdot Q(s_t, a_t) + \alpha \cdot \overbrace{\left(r_t + \gamma \cdot \max_a Q(s_{t+1}, a)\right)}^{\text{learned value}} \tag{2.2}$$

where r_t denotes the reward at timestep t, α denotes the learning rate, and γ is a discount factor. This update step $Q(s_t, a_t)$ denotes the expected reward over time, with a discount factor which diminishes the value of distant rewards over proximal ones. While straightforward to calculate, this model requires a large storage table and fails to assign accurate predictions to unseen state-action pairs.

Although actions are typically chosen to maximize the Q-function, we also implement an exploration-exploitation tradeoff with ϵ-scheduling. Under this model, we choose a random action with probability ϵ and the argmax of the Q-function with probability $(1-\epsilon)$. Over many iterations, ϵ gradually decreases. Roughly, this prevents the model from getting stuck in local maxima before it has sufficiently explored the region of potential policies.

2.1.2.2 Tabular Q-Learning (with Monte Carlo Estimation)

Monte Carlo methods allow us to apply a probabilistic sampling approach to tabular Q-learning. Under this model, which was applied to *Task & Talk* in Das et al. (2017), Q-values for specific state-action pairs are calculated by simulating potential continuations of task iterations and averaging rewards across these. That is, instead of directly choosing actions with the highest Q-value, we simulate $K = 100$ possible dialogue continuations and average reward as follows (Sutton and Barto, 2018):

$$Q(s_t, a_t) = \frac{1}{K} \sum_{t=1}^{K} R_t \tag{2.3}$$

where R_k represents the return of a single simulation. Given this Q-function, we sample actions $a' = \text{argmax}\, Q(s_t, a)$. Because each Q-function update is more accurate than in standard tabular Q-learning, the total number of iterations before convergence decreases.

2.1.2.3 REINFORCE

REINFORCE (Williams, 1992) is a Monte Carlo policy gradient algorithm which generates probability distributions over actions given states, denoted $\pi_\theta(a_t \mid s_t)$. This model is based on the policy gradient theorem, which equates the derivative of expected reward to the expectation of the reward and the log policy gradient:

$$\nabla \mathbb{E}_{\pi_\theta}[r(\tau)] = \left[r(\tau) \left(\sum_{t=1}^{T} \nabla \log \pi_\theta(a_t \mid s_t) \right) \right] \tag{2.4}$$

where τ denotes a series of actions resulting in a reward $r(\tau)$. Kottur et al. (2017) applies this model to several variations of the *Task & Talk* reference game, and we will use it as the starting point for our pragmatic model, which is described in the following section.

2.1.2.4 Pragmatic REINFORCE

Recall the formalization of computational pragmatics in Section 1.1.3. Here, we modify the incremental version of this framework to fit a reinforcement learning paradigm. Using the standard REINFORCE algorithm with hyperparameters from Kottur et al. (2017), we train stochastic policies $\pi_Q(\hat{w} \mid s_t^Q; \theta_Q)$ and $\pi_A(a_t \mid s_t^A; \theta_A)$ and use them as base agents in our pragmatic calculation:

$$P\left(a_t \mid s_t^A\right) \propto \pi_A\left(a_t \mid s_t^A; \theta_A\right) \cdot \mathbb{E}_{s_0^Q}\left[\pi_Q\left(\hat{w} \mid s_0^Q; \theta_Q\right) \cdot U(\hat{w})\right] \tag{2.5}$$

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. We define $U(\hat{w}) = 1$ if true *or* partially true, and $U(\hat{w}) = 0$ otherwise: i.e. if $\hat{w}$ = *(red)* and the target is *(red, square)*, $U(\hat{w}) = 1$. Note that, for the sake of implementation, we use Q-BOT's policy π_Q in the model $P\left(a_t \mid s_t^A\right)$ 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.

2.1.3 Modified *Task & Talk*

We identify various issues with the original *Task & Talk* framework which are detailed in Tomlin and Pavlick (2018). Among these, we note that Q-BOT *must* evenly partition the search space for optimal communication to occur. We therefore consider several modifications to *Task & Talk*, summarized in Table 2.1. Below, we describe the considered task variants.

2.1.3.1 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.

	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 $\|V_A\|$	4	4	8

Table 2.1: Summary of *Task & Talk* modifications.

(2017) restricts A-BOT's state $s_A^t = [I, q_t]$ to only include the most recent utterance from Q-BOT. We 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. However, we choose to ignore this variant because it reduces to an optimization problem by technicality: marginally increasing the vocabulary size or modifying the referent set would immediately break this strategy, which is merely an artifact of highly-specified task parameters (Tomlin and Pavlick, 2018).

2.1.3.2 Simplified 4x4 Baseline

Consider a simplified version of this task, in which A-BOT has access to the task specification G. Because of this, Q-BOT's utterances may be functionally ignored, and 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.

Note that this 4x4 Baseline is susceptible to partitioning strategies, which are discussed in Kottur et al. (2017) and visualized in Figure 2.2. These strategies occur when compositional and non-compositional mappings are equally optimal at every iteration, i.e., when any two partitioning strategies are homomorphic. We must consider additional task modifications to avoid such partitioning strategies.

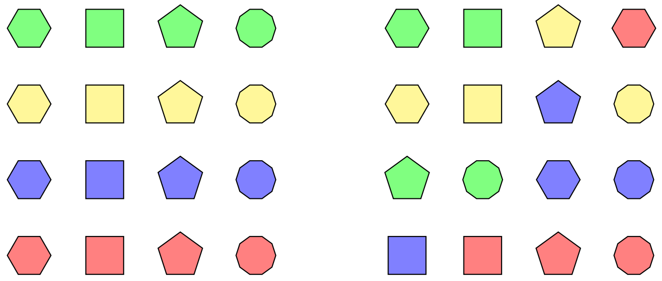


Figure 2.2: Two possible partitioning strategies in the 4x4 variant. Let rows denote tokens $\{a_1\}$ and columns denote tokens $\{a_2\}$. While both partitioning strategies are equally optimal, only the left configuration is *grounded*.

2.1.3.3 Modified 4x4 Multitask

While the communication protocol from Section 2.1.3.1 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 4x4 setting discussed in Section 2.1.3.2 above. While we observe that grounded communication does not emerge in the 4x4 Baseline due to partitioning strategies, 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. Finally, we use a curriculum learning method in which one-attribute tasks are presented before two-attribute tasks, allowing Q-BOT to develop a policy for the one-attribute case. These modifications are summarized in Table 2.1.

2.1.4 Model Architecture

We model Q-BOT and A-BOT using tools from deep reinforcement learning (Li, 2017). In particular, both agents have input and output networks which are parametrized in terms of long short-term memory networks, or LSTMs (Hochreiter and Schmidhuber, 1997). These networks input discrete-valued encodings of tasks and referents into real-valued policies which may be optimized via deep learning. The structure of these networks is shown in Figure 2.3, and modified accordingly for the task variants discussed in Section 2.1.3.

2.2 Model Results

We develop computational experiments corresponding to each of the mechanisms listed in Box 1, with the exception of iterated transmission effects. We exclude this

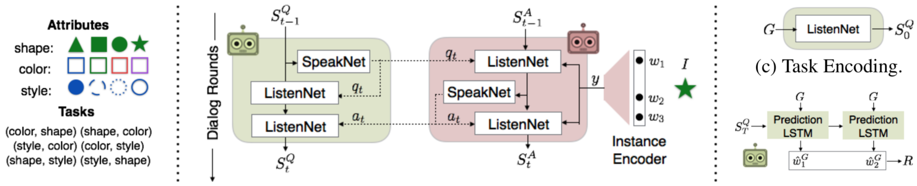


Figure 2.3: Model architecture diagram reproduced from Kottur et al. (2017).

mechanism because it has been well-supported in both computational simulations and human experiments. However, as stated, we expect that various other mechanisms may lead to compositional referring behavior without the reliance on a large population of speakers.

2.2.1 Noisy Channel Model

We will briefly discuss preliminary results concerning the effect of noise on the compositionality of emergent communication protocols. These results are restricted to the tabular Q-learning settings. We consider all possible combinations of the following modifications:

- Noisy replacement of single tokens with randomly chosen alternatives;
- Addition of partial rewards for partially correct guesses;
- Implementation of incremental sequence samplers (Futrell, 2017).

While the rate at which tokens are noisily modified determines the convergence rate of the model, none of these features was shown to have a significant effect on the groundedness score. Note that these results are inconclusive, and a much more thorough investigation would be required to enumerate all possible integrations of a noisy channel model across all reinforcement learning settings. However, they are suggestive that noise alone cannot contribute to compositional referring behavior in the *Task & Talk* reference game.

2.2.2 Pragmatic Model Achieves Compositionality

We run each of the reinforcement learning algorithms discussed in Section 2.1.2 for 100 trials and average groundedness scores post-convergence. All models except Pragmatic REINFORCE receive groundedness scores at chance, as shown in Table 2.2. This provides strong evidence that the incremental pragmatic model may be an adequate mechanism for compositionality, but with the caveat that this model only achieves high groundedness scores on the 4x4 Multitask variant. In the following sections, we will briefly ablate key distinctions between the 4x4 Baseline and 4x4 Multitask settings.

	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 2.2: Mean policy groundedness scores (Equation 1.2) 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.

2.2.2.1 Effect of Task Knowledge

Recall that Q-BOT does not observe the task specification G in the 4x4 Multitask setting. We argue that this is a necessary factor to achieving compositional behavior. In particular, when the task G is observed, each of the three tasks may be treated separately by the model, with the *(shape, color)* task reducing to the 4x4 Baseline.

	4x4 Multitask	Task Knowledge
Tabular Q-Learning	0.181	0.183
Tabular Q-Learning (MC)	0.182	0.182
REINFORCE	0.188	0.181
Pragmatic REINFORCE	0.874	0.181

Table 2.3: Mean policy groundedness scores, over 20 iterations in the *task knowledge* setting, which is characterized by Q-BOT observing G.

2.2.2.2 Effect of Partial Rewards

Similarly, we argue that the utility function $U(\hat{w})$ used in Pragmatic REINFORCE is required for compositionality in *Task & Talk*. We modify the utility function to require exact equivalence, i.e., if $\hat{w}$ = *(red)* and the target is *(red, square)*, $U(\hat{w}) = 0$ in this modified setting. Indeed, running 20 iterations gives a mean groundedness score of 0.185 under this condition, which is at chance.

2.2.2.3 Effect of Curriculum Learning

Finally, recall that one-attribute tasks are presented sequentially before two-attribute tasks in every iteration of the 4x4 Multitask *Task & Talk*. Note that it is immediately possible for the model to achieve perfect grounded communication in the one-attribute case. In the pragmatic model, therefore, the two-token referring expressions are largely influenced by existing single-attribute mappings. We define the probability of choosing a single-attribute task as follows:

$$P(\text{single-attribute}) = g_0 \left[(1-\lambda) \cdot \frac{\#\ \text{single-attribute}}{\#\ \text{two-attribute}} + \lambda \right] \tag{2.6}$$

where g_0 is a boolean value representing whether the model has achieved perfect accuracy on single-attribute mappings and λ is a hyperparameter which determines the extent of curriculum learning. Therefore, when $\lambda = g_0 = 0$, the model will randomly choose tasks and referents without respect to the curriculum. We show the effects of curriculum learning in Figure 2.4 for several values of λ and note that it has a significant effect on groundedness.

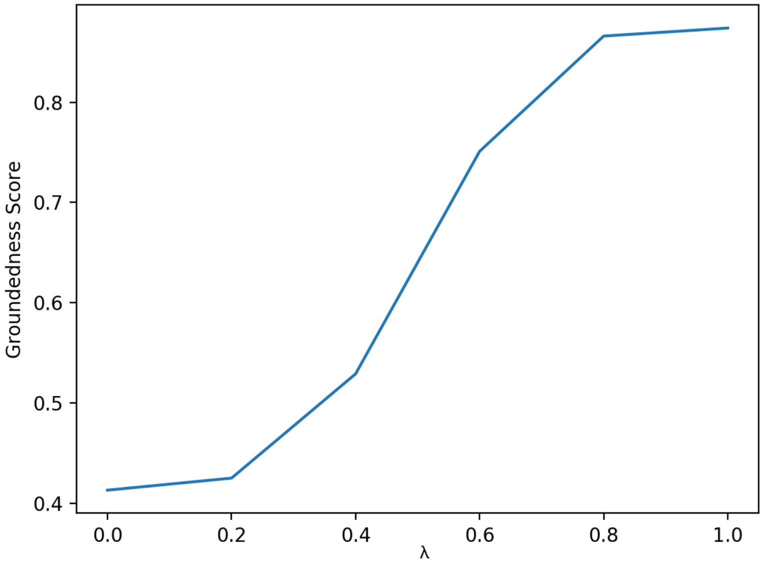


Figure 2.4: Mean groundedness scores for $\lambda \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ where λ parametrizes the extent of curriculum learning as defined in Equation (2.6). Therefore, although there is a significant effect of curriculum learning on compositionality, it is not solely responsible for the high groundedness score.

2.2.3 Compression in Vanilla RNNs vs. LSTMs

Compression effects occur in humans when it becomes difficult to memorize an increasingly large set of non-compositional mappings (Yang, 2016), but the computational analogue is unclear. Tabular methods, for example, are capable of storing arbitrarily large policies and are constrained only by computer memory. One potentially corresponding notion is the distinction between recurrent neural networks (RNNs) and LSTMs. While LSTMs are a subtype of RNN and both are used as language-modeling tools, LSTMs contain memory cells and are typically preferred over vanilla RNNs for memory-intensive tasks. Because of this, LSTMs are more commonly used in natural language processing. However, we note that LSTMs actually receive *lower* groundedness scores on *Task & Talk*, as shown in Table 2.4.

	4x4 Baseline	4x4 Multitask
(RNN) REINFORCE	0.203	0.254
(RNN) Pragmatic REINFORCE	0.210	0.890
(LSTM) REINFORCE	0.150	0.188
(LSTM) Pragmatic REINFORCE	0.153	0.874

Table 2.4: Mean policy groundedness scores (Equation 1.2) across 20 iterations. LSTM data comes from Table 2.2 and is repeated here for comparison. Disclaimer: not all iterations in the RNN settings converge to perfect accuracy on *Task & Talk*, so we calculate groundedness scores only for policies which successfully account for referents in the training data.

These results are consistent with the notion that compression effects are related to memory strength. However, it is difficult to draw conclusions about the effects of memory and compression in humans from such data. We will therefore now turn to our iterated learning experiments on human subjects.

Chapter 3

Human Experiments

To evaluate the extensibility of our models and address computationally imprecise mechanisms such as compression, we develop an experimental counterpart to *Task & Talk* for evaluation on human participants. Certain modifications to *Task & Talk*, such as reducing the number of referents, must be made in order to ensure strong human performance on this task.

3.1 Experimental Design

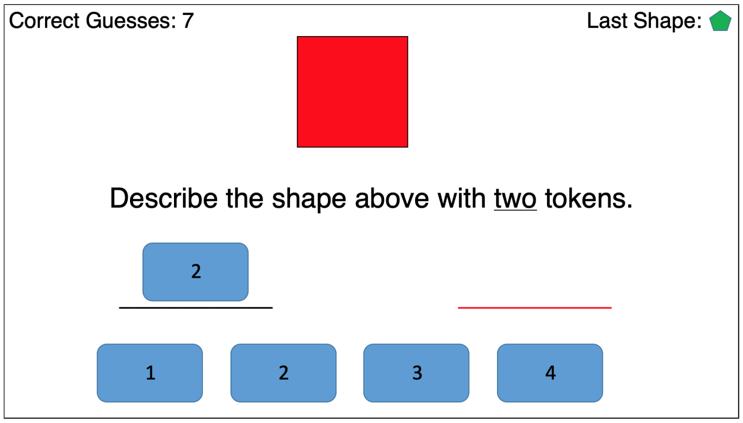
3.1.1 Task Specification

We develop an iterated reference game task between two participants, a SENDER and RECEIVER, corresponding to A-BOT and Q-BOT respectively. As depicted in Figure 3.1, the sender is shown a single referent and prompted to describe it with a sequence of tokens. The receiver sees the signal and is prompted to guess a single item from the referent set. Both players receive binary feedback after the guessing phase, and this game repeats with randomly chosen referents until the players make 10 correct guesses in a row.

The referent set varies across the experiments as described in Section 3.2.

3.1.2 Method and Participants

We recruited 20 undergraduate and graduate students from Brown University to participate in the experiments described below. Participants were divided into sender-receiver pairs and split across the two experimental conditions. Each subject participated in only a single experiment, consisting of an iterated signaling game repeated until perfect accuracy was achieved. Trials lasted from 5-20 minutes.



Correct Guesses: 7

Last Shape:

Describe the shape above with two tokens.

Figure 3.1: Proposed graphical interface for sender

The reference game itself was implemented as a website application written in *Node.js*. Trials were hosted over a single server with unique URLs assigned to senders and receivers, who viewed the task from their own laptops. Players were unable to see each other's screens or communicate with one another outside of the signaling interface. Player data, including a full history of referents and referring expressions, was logged onto the server.

3.1.3 Challenges with Human Experiments

We identify numerous implementational and analytical challenges with these human experiments which are summarized in Box 2. As discussed in Section 3.2.1, planning poses a particularly strong issue in Experiment I. This occurs because senders can quickly infer the entire referent set and develop a compositional communication policy which accounts for all predicted referents; when this happens, receivers have no input in the communication policy, and compression effects cannot be observed because communication policies do not evolve over time.

Another challenge involves the difficulty of the signaling task. In an initial pilot study, participants failed to complete the 4x4 Baseline task described in Section 2.1.3.2. With too few referents, however, participants can easily memorize every referring expression and face minimal compression effects. We eventually settled on a 9-referent set for both experiments, with most subjects citing difficulty but with all eventually completing the task.

Finally, we note that this experiment encounters many of the typical issues faced by artificial language learning tasks. In particular: because human subjects have existing vocabularies and language for describing these referents, we expect a confounding effect. For example, knowing that compositional referring expressions exist in one language may bias humans toward compositional strategies in the reference

Box 2: Challenges with Human Experiment

Various factors cause difficulties in the human version of *Task & Talk*. The most pertinent of these are summarized below:

- Planning effects – because senders roughly know the reference space, they may preemptively plan compositional strategies without input from the receiver.
- Balancing number of referents – participants should encounter memory constraints while still being able to successfully complete the task
- Unnaturalness – as with other types of artificial language learning experiments, human participants are influenced by their existing vocabulary and language behavior

Computational experiments in emergent communication face their own challenges, but combining these two methods may help address uncertainties that occur in one or the other.

game. An additional source of unnaturalness stems from the signaling interface: because our work relies on having a fixed vocabulary size and does not involve starter referring expressions, we cannot mirror open-domain paradigms such as Kirby et al. (2014) which allow participants to generate morphological expressions.

3.2 Experimental Results

We report the results of two experiments, which differ only in the set of target referents used. Due to the number of participants and the time required to run additional trials, the provided analysis is primarily qualitative.

3.2.1 Experiment I: Dense State Space

We begin with a 3x3 state space, visualized in Figure 3.2. Senders and receivers communicate with a Simon board (Proctor, 2011). This paradigm was chosen specifically to avoid symbolic mappings which may occur in systems with numerical or character-based tokens, e.g., associating triangles to token "3." We holdout two referents for evaluation post-convergence; that is, once participants have made 10 successful guesses in a row, they are shown an unseen referent and tasked with correctly signaling it. This is intended to measure the degree to which the invented communication policy generalizes and is strongly correlated with compositionality.

Out of the five pairs of participants run on this trial, four of them immediately resorted to a planned compositional strategy. In each of these cases, the sender rarely

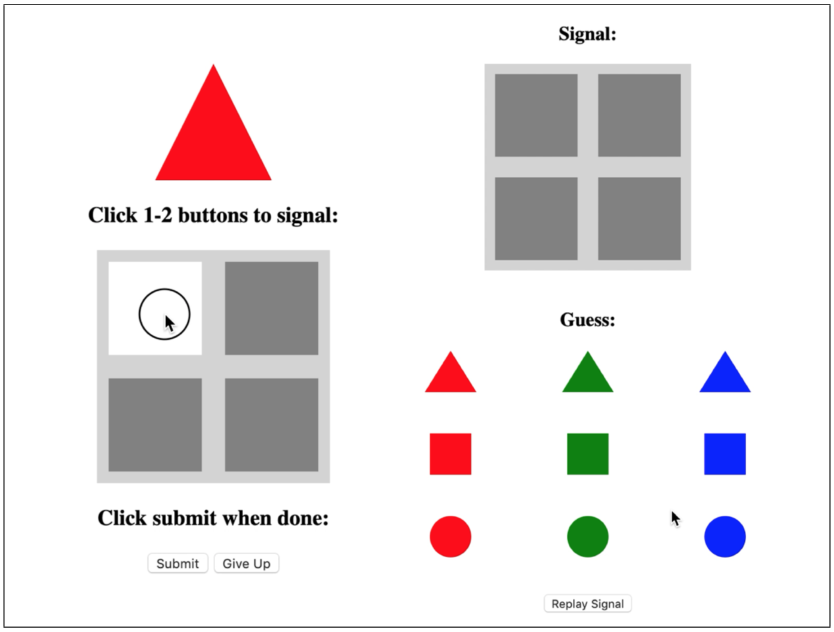


Figure 3.2: Final graphical interface for Experiment I

(if ever) deviated from their initial strategy, and the receiver was forced to guess and adapt to the planned protocol. In the remaining fifth trial, the sender was again perfectly consistent, but the chosen strategy was non-compositional. Only participants in the first four trials were able to succesfully signal the heldout referent. The learning curves associated with these trials are shown in Figure 3.5.

3.2.2 Experiment II: Sparse State Space

Due to the role of planning in Experiment I, it is impossible to pinpoint specific mechanisms leading to compositionality. We therefore consider a sparser state space with the goal of undermining planning strategies. In this version, shown in Figure 3.3, we replace our 3x3 state space of shapes with a 3x3x3 state space of avatars. While we maintain the same number of referents, they are sparsely chosen from this space, making it difficult for the sender to implement a planning strategy.

In this revised experiment, we hypothesized that compression effects would be realized as a U-shaped learning curve, as in Figure 3.4. This is expected to occur because senders choose an initial communication strategy which is later revealed to be non-optimal in terms of required memory. Indeed, as shown in Figure 3.5, the learning curves for this experiment do exhibit a slight U-shaped dip before converging to a solution. Groundedness typically increases after these compression periods.

3.3 Analysis of Results

We observe preliminary evidence of compression effects in Experiment II. This is typically characterized by a simultaneous decline in accuracy and increase in groundedness shortly before convergence to an optimal strategy. Despite this, we note that the groundedness of policies is far above chance prior to compression. Because of this, we must assume that compositionality in this setting is the result of multiple mechanisms, one of which is compression.

While this methodology does not allow us to directly pinpoint the other contributing mechanisms, we note that this effect is consistent with the incremental pragmatic model of Section 2.2.2. Future work investigating this mechanism in humans might attempt to recreate computational results related to the pragmatic mechanism, such as effects of curriculum learning.

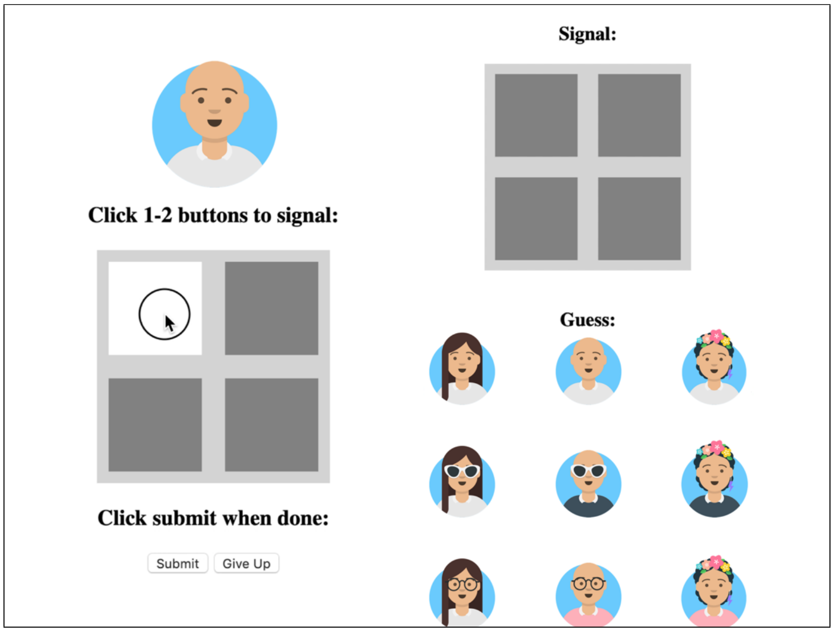


Figure 3.3: Final graphical interface for Experiment II

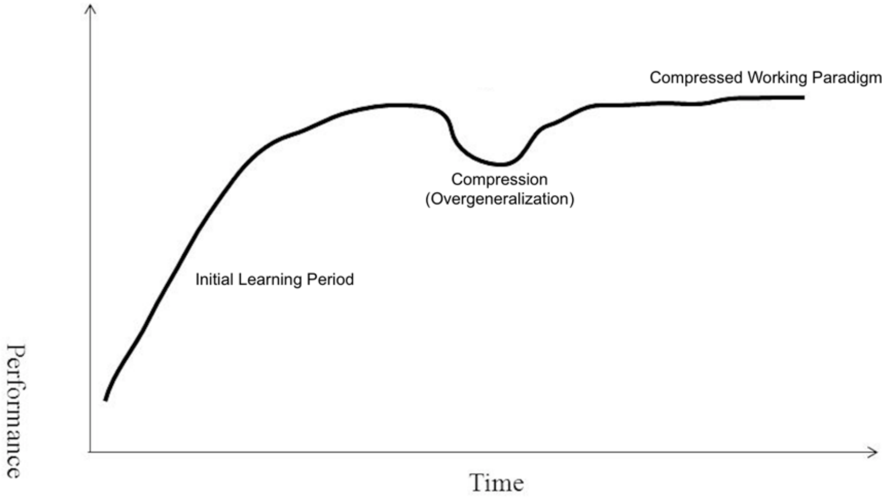


Figure 3.4: Idealized U-shaped learning curve associated with compression model. Compression is expected to temporarily affect performance due to confusion associated with re-mapping and overgeneralization effects.

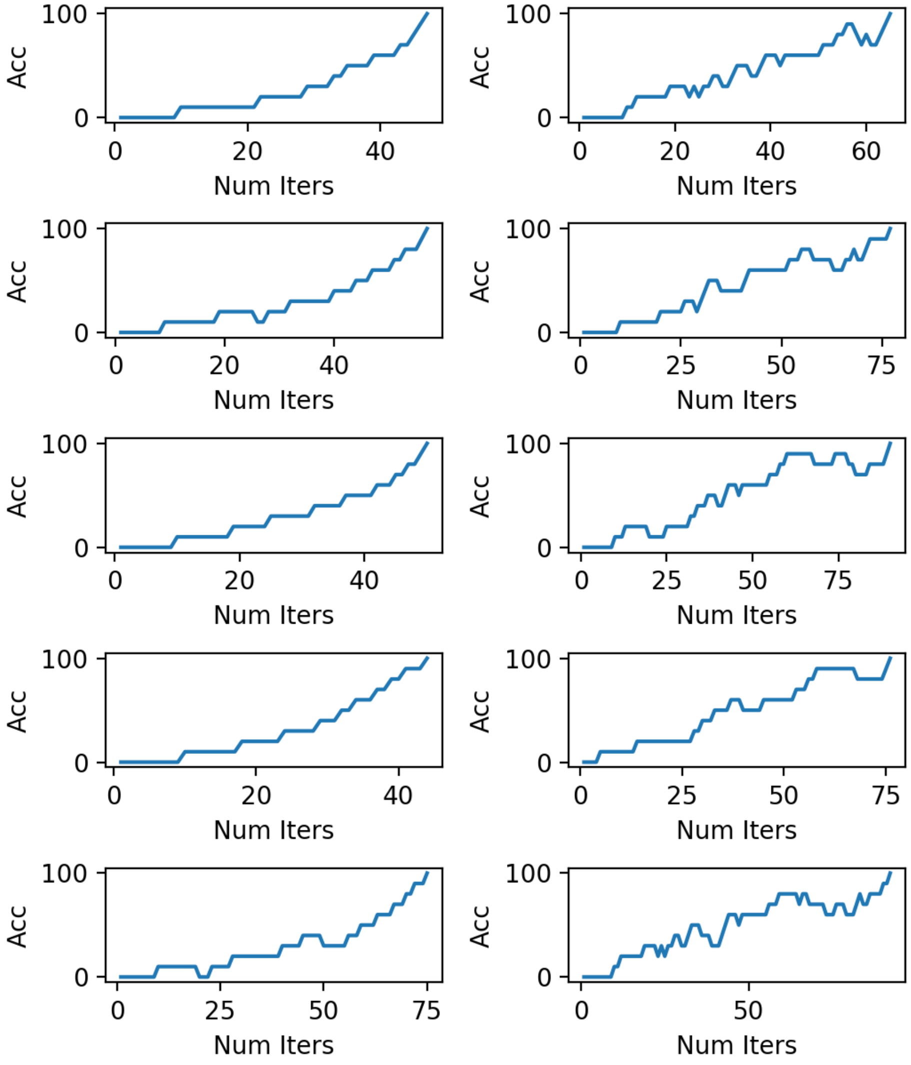


Figure 3.5: Learning curves for Experiment I (left) and Experiment II (right) over 20 participants. In all graphs, the x-axis represents the number of iterations and the y-axis represents the accuracy over the past 10 trials.

Chapter 4

Discussion

We considered computational and human experiments across several variations of the *Task & Talk* reference game. Although emergent communication and iterated learning both struggle to make definitive claims about the emergence of linguistic structure, they provide powerful suggestive evidence when combined. We consider the ramifications of these experiments for mechanisms leading to compositionality in Section 4.1.

Seeing positive evidence for the pragmatic model, we then briefly outline its motivations in Section 4.2. We consider its extensibility to other linguistic phenomena in Section 4.3 and discuss challenges faced in applying similar computational approaches to such domains. Finally, in Section 4.4, we take an optimistic look at how emergent communication may influence the future of linguistic research.

4.1 Reevaluating Mechanisms of Compositionality

Recall the potential mechanisms leading to compositionality listed in Box 1, and recall that these are neither necessarily contradictory nor comprehensive. With these caveats in mind, we will briefly return to each of the mechanisms and evaluate their respective potentials as mechanisms for compositional referring behavior.

First, as noted previously, iterated transmission effects have already shown sufficient promise in simulations and human experiments (Kirby et al., 2014, 2015). Second, we found insufficient evidence for the effect of noisy-channel models in the *Task & Talk* setting; however, this result is preliminary due to the numerous potential ways to integrate noisy-channel effects. Third, we find evidence for the compression model in humans, although its effect is small and the computational analogue unclear. Finally, we find strong computational evidence for the pragmatic model; this model is also consistent with the human data.

While other mechanisms might also exist, a reasonable hypothesis would be that some combination of pragmatic reasoning and compression are responsible for compositionality in Experiment II of the human studies.

4.2 Feasibility of Pragmatic Mechanism

We will briefly motivate why incremental pragmatics may be a promising mechanism for compositionality. Note first that the classical RSA model from Section 1.1.3.2 weakly enforces injectivity between utterances and their meanings. That is, RSA will typically prevent two utterances from being mapped to the exact same meaning unless a strong prior on the base agent suggests otherwise.

Meanwhile, because incremental pragmatics generates alternatives at the word-level rather than the sentence-level, it may weakly enforce an injective mapping between word tokens and their meanings. When meanings correspond to referent attributes, as in the settings described above, incremental pragmatics may therefore be responsible for emergent groundedness.

4.3 Extensions to More Complex Linguistic Phenomena

While this work has focused on the compositionality of referring expressions, which might correspond to adjective-noun pairings in natural language, it is not unreasonable to suggest that the mechanisms presented here might be expanded to account for more complex linguistic phenomena. To sketch a specific example, the incremental pragmatic model might be applied to explain relationships between argument structure constructions. In this setting, we might consider arguments to correspond to the notion of referent attributes. We could then, for example, predict that separate terms *eat* and *devour* emerge in order to maximize informativity while balancing pressures related to memory and consistency in transmission.

However, studying such complex phenomena with emergent communication poses a new set of technical challenges. First, while the models presented here are strictly non-recursive, they presumably must be integrated into a PCFG or similar in order to account for various syntactic and semantic structures. Further, while deep learning has led to significant progress in natural language processing, reinforcement learning problems such as knowledge representation and concept-learning are still computationally difficult and prerequisite to studying certain types of language behavior with emergent communication.

4.4 New Directions for Experiments in Artificial Language Learning

Addressing the challenges at the intersection of language and reinforcement learning could lead to vibrant new approaches to artificial language learning. While we have seen that emergent communication systems in general do not exhibit human-like behavior, it is possible to significantly modify their behavior by adding certain inductive biases. With the correct biases, these emergent communication systems provide a promising alternative to research in artificial language learning.

Human studies in artificial language learning are ripe with challenges. In particular, such experiments are often costly, time-intensive, and limited in scale. Furthermore, humans are typically biased by their existing linguistic knowledge. Reinforcement learning agents avoid all of these issues and offer the potential for truly scalable artificial language learning experiments. While the complexity of linguistic phenomena studied in these experiments is usually limited by what humans can acquire over a short timeframe, emergent communication systems have the potential to exhibit much richer language behavior if they contain the right inductive biases to do so. In the mean time, we expect that iterated learning and emergent communication will continue to go hand-in-hand as complementary research paradigms.

Bibliography

Andreas, J., Dragan, A., and Klein, D. (2017). Translating neuralese. *arXiv preprint arXiv:1704.06960*.

Andreas, J. and Klein, D. (2016). Reasoning about pragmatics with neural listeners and speakers. *arXiv preprint arXiv:1604.00562*.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, 59:617–645.

Bergen, L., Levy, R., and Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial life*, 8(1):25–54.

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

Christiansen, M. H., Chater, N., and Culicover, P. W. (2016). *Creating language: Integrating evolution, acquisition, and processing*. MIT Press.

Cohn-Gordon, R., Goodman, N., and Potts, C. (2018a). Pragmatically informative image captioning with character-level reference. *arXiv preprint arXiv:1804.05417*.

Cohn-Gordon, R., Goodman, N. D., and Potts, C. (2018b). 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. *arXiv preprint arXiv:1703.06585*.

De Beule, J. and Bergen, B. K. (2006). On the emergence of compositionality. In *The Evolution Of Language*, pages 35–42. World Scientific.

Foerster, J., Assael, I. A., de Freitas, N., and Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 2137–2145.

Foerster, J., Chen, R. Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., and Mordatch, I. (2018). Learning with opponent-learning awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 122–130. International Foundation for Autonomous Agents and Multiagent Systems.

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

Frege, G. (1892). Uber sinn und bedeutung. *Zeitschrift fur Philosophie und philosophische Kritik*, 100(1):25–50.

Friederici, A. D. and Chomsky, N. (2017). *Language in Our Brain: The Origins of a Uniquely Human Capacity*. MIT Press.

Futrell, R. (2017). *Memory and locality in natural language*. PhD thesis, PhD thesis. Cambridge, MA: Massachusetts Institute of Technology.

Gibson, E., Bergen, L., and Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20):8051–8056.

Gibson, E., Futrell, R., Piantadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., and Levy, R. P. (2019). How efficiency shapes human language, tics 2019.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

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

Grice, H. P. (1975). Logic and conversation. *1975*, pages 41–58.

Hawkins, R. X., Frank, M., and Goodman, N. D. (2017). Convention-formation in iterated reference games. In *CogSci*.

Hermann, K. M., Hill, F., Green, S., Wang, F., Faulkner, R., Soyer, H., Szepesvari, D., Czarnecki, W., Jaderberg, M., Teplyashin, D., et al. (2017). Grounded language learning in a simulated 3d world. *arXiv preprint arXiv:1706.06551*.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

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

Kao, J., Bergen, L., and Goodman, N. (2014a). Formalizing the pragmatics of metaphor understanding. In *Proceedings of the annual meeting of the Cognitive Science Society*, volume 36.

Kao, J. T., Wu, J. Y., Bergen, L., and Goodman, N. D. (2014b). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33):12002–12007.

Kirby, S., Griffiths, T., and Smith, K. (2014). Iterated learning and the evolution of language. *Current opinion in neurobiology*, 28:108–114.

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.

Kottur, S., Moura, J. M., Lee, S., and Batra, D. (2017). Natural language does not emerge'naturally'in multi-agent dialog. *arXiv preprint arXiv:1706.08502*.

Krauss, R. M. and Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1(1-12):113–114.

Levy, R. P. (2018). Communicative efficiency, uniform information density, and the rational speech act theory.

Li, Y. (2017). Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*.

Linzen, T. (2019). What can linguistics and deep learning contribute to each other? response to pater. *Language*.

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, O. P., and Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems*, pages 6382–6393.

Mikolov, T., Joulin, A., and Baroni, M. (2016). A roadmap towards machine intelligence. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 29–61. Springer.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529.

Monroe, W. and Potts, C. (2015). Learning in the rational speech acts model. *arXiv preprint arXiv:1510.06807*.

Mordatch, I. and Abbeel, P. (2017). Emergence of grounded compositional language in multi-agent populations. *arXiv preprint arXiv:1703.04908*.

Nowak, M. A. and Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033.

O'Grady, W. (2008). The emergentist program. *Lingua*, 118(4):447–464.

Partee, B. (1984). Compositionality. *Varieties of formal semantics*, 3:281–311.

Pater, J. (2019). Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language*.

Potts, C. (2019). A case for deep learning in semantics: Response to pater. *Language*.

Potts, C., Lassiter, D., Levy, R., and Frank, M. C. (2016). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, 33(4):755–802.

Proctor, R. W. (2011). Playing the simon game: Use of the simon task for investigating human information processing. *Acta Psychologica*, 136(2):182–188.

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.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

Smith, K., Brighton, H., and Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Advances in Complex Systems*, 6(04):537–558.

Sukhbaatar, S., Fergus, R., et al. (2016). Learning multiagent communication with backpropagation. In *Advances in Neural Information Processing Systems*, pages 2244–2252.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tomlin, N. and Pavlick, E. (2018). Incremental pragmatics and emergent communication. *Neural Information Processing Systems Workshop on Emergent Communication*.

Tomlin, N. and Pavlick, E. (2019). Emergent compositionality in signaling games. *Forty-First Annual Conference of the Cognitive Science Society*.

Wagner, K., Reggia, J. A., Uriagereka, J., and Wilkinson, G. S. (2003). Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1):37–69.

Wang, S. I., Liang, P., and Manning, C. D. (2016). Learning language games through interaction. *arXiv preprint arXiv:1606.02447*.

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

Yang, C. (2016). The price of linguistic productivity.