Learning Language: An Experiment

Daniel Houser and Yang Yang

November 2020

Discussion Paper

Interdisciplinary Center for Economic Science
4400 University Drive, MSN 1B2, Fairfax, VA 22030
Tel: +1-703-993-4850 Fax: +1-703-993-4851
ICES Website: http://ices.gmu.edu
ICES RePEc Archive Online at: http://edirc.repec.org/data/icgmuus.html

Learning Language: An Experiment

Daniel Houser [a] *Yang Yang* [b1]

[a] Interdisciplinary Center for Economic Science, George Mason University, 4400 University Drive, Fairfax, VA 22030, United States of America. Email: dhouser@gmu.edu

[b] Lingnan College, Sun Yat-sen University, 135 Xingang West Rd, Guangzhou, 510275, China. Email: yangyang8@mail.sysu.edu.cn

Abstract

We develop a method for random assignment of language to participants in a controlled laboratory experiment, and use this to test the hypothesis that languages are learned more quickly when they can be identified with fewer number of observations. While the theory based on this hypothesis has generated substantial attention since being advanced by Blume (2005), evidence on its empirical validity has been elusive. Here we develop a novel extension of coordination games within which languages emerge endogenously. We show, first, that one can control features of an emergent language by varying the game's incentives. This enables us to compare speed of learning across participants randomly assigned to different languages. Our data provide cogent evidence supporting the above hypothesis and Blume's (2005) theory: Languages with compositional structures can be identified with fewer observations and are learned more quickly, and in this sense are efficient. Despite this, we find inefficient languages to sometimes emerge when they can be expressed using simple rules.

Keywords: testing the efficiency theory of language, random assignment of language, laboratory experiment

JEL codes: C91, D83

[1] Corresponding author.

1. Introduction

We develop an approach for random assignment of language to participants in a laboratory experiment, and use this to test the hypothesis that languages are learned more quickly when they are determined by fewer observations.[2] This hypothesis is the underlying premise of the Blume's (2005) theory that has generated substantial multi-disciplinary attention[3]. Although several papers have made efforts to test this hypothesis, rigorous empirical evidence on its validity has been slow to emerge. Our approach allows us to compare learning efficiency between treatments with languages that require different number of observations to be determined, in the sense of Blume (2005). Our data provide rigorous evidence that languages are learned more quickly when they require fewer number of observations to be determined.

Blume (2005) hypothesizes that people learn language by observation (or example), so that languages that require fewer examples to determine should be learned more quickly. He then proves that, under regularity conditions, the language requiring the fewest examples to describe a given set of objects is compositional. When combined with the hypothesis that people learn by example, this provides an explanation for why all human languages have a compositional structure: they can be learned more easily than non-compositional languages.

Whether languages that require fewer examples to determine can in fact be learned more easily is of course an empirical question. One early paper to study this is Hong et al. (2017). Using a sender-receiver laboratory game, Hong et al. (2017) demonstrated convincingly that compositional grammars do emerge in the lab. While compositional languages often emerged, they also found non-compositional language to emerge, and they found inconsistent results regarding the number of new observations with which these different types of language were learned. One reason for this might be that they compare speed of learning languages that emerged within a single treatment, leaving open the possibility of selection effects: people who discover and use a compositional language may differ from those who discover and use

[2] The way observations are used to identify languages will be detailed below. Intuitively, a language is "identified" when it is known how to use it to describe objects. Different language structures require different numbers of observations (or examples) of language use to identify the language.

[3] See e.g., Hernández et al. (2012), Kyun (2012), Rai, et al. (2012), Suzuki (2020). Relatedly, from a different perspective, Blume (2000) points out the usefulness of a language's structure in indicating new objects, and describes optimal learning with partial language in coordination games. The latter was tested by Blume and Gneezy (2000) using a laboratory experiment.

alternative languages, which may confound the within-treatment comparison due to (observable or unobservable) differences between subjects who self-select into different languages. Our paper complements the important initial effort by Hong et al (2017) by developing a design that allows for between-treatment randomized comparisons. We show this results in cogent evidence regarding speed of learning different types of languages.

Our experiment builds from the coordination game introduced by Selten and Warglien (2007).[4] Our participants play a communication game in pairs. Each pair is asked to label objects (emojis) using different pre-determined symbols (fruits). If their fruit labels match they can earn money, and if not they do not. The crucial feature of our design is that we vary the costs of using the fruit symbols between treatments. We show that doing this affects in predictable ways the nature of the emergent language. In particular, by varying incentives we can promote the endogenous emergence of two different compositional languages and one non-compositional language, all of which describe the same set of emoji objects.[5] Moreover, these languages differ in the number of observations required to learn them, with one of the compositional languages requiring the least, and the non-compositional language requiring the most.

We find, as did Selten and Warglein (2007), Hong et al. (2017) and Hong and Zhao (2017), that languages do emerge in all of our treatments and that some of these languages are compositional.[6] The costs of the fruit symbols in one of our treatments promotes a compositional language that requires only three observations to learn (we denote this the 3-3 treatment), while our other treatment's symbol costs promote a compositional language that requires four observations to learn, or alternatively a non-compositional language that requires five observations (we denote this the 4-2 treatment). Our main result is that the speed of learning

[4] A rich literature investigates the emergence of language and meaning, or the impact of meaning in economic contexts. See, e.g., Chan et al. (2011), Cremer et al. (2007), Devetag (2005), Franke (2014, 2016), Galantucci and Garrod (2010, 2011), and Weber and Camerer (2003).

[5] Rubinstein (1996) was the first to provide a formal economic analysis of the way incentives ("evolutionary forces") of a language environment can impact the emergence and development of language. More broadly, work on economics and language includes Lipman (2003), Marschak (1965), and Rubinstein (2000).

[6] Compositional grammars, including all human grammars, have the advantage that they allow speakers to describe objects never before seen. This differs, for example, from simple codes which only allow one to describe known objects. A large literature highlights and explores the unique advantages of compositional grammars. See e.g., Bresnan (1982), Steinert-Threlkeld (2016) or Szabo (2013).

(measured as number of matches successfully created by a given round of the game) is economically and statistically significantly higher in the 3-3 treatment, which includes incentives that promote the language requiring the fewest number of observations to learn.

The remainder of the paper is organized as follows. In section 2 we comment on the related literature. Section 3 reviews the model of language suggested by Blume (2005), and uses it to motivate our experiment design and hypotheses, which we report in section 4. Section 5 discusses our results, and section 6 concludes.

2. Literature Review

Selten and Warglein (2007) (and earlier Blume, 1998) study the emergence of language in the laboratory. Selten and Warglein (2007) developed a communication game, a variant of which we use in our study, to investigate the structure of emergent language. They were particularly interested in the structure of emergent language, and in particular whether it is compositional. They found that perfect compositional structures emerged relatively infrequently, a result that we share. Our paper differs from theirs, however, in our focus on the ease and speed with which different languages can be learned.

Hong et al. (2017) and Hong and Zhao (2017) report data from a sender-receiver game designed to study how language emerges in the lab. Both papers focus on the structure of language, with Hong and Zhao (2017) asking how the set of objects impacts the emergent language when a limited set of symbols can be used in expressions, while Hong et al. (2017) focus on whether emergent language is compositional, and whether compositional grammars can be learned more quickly than other emergent languages. Our paper is closely aligned to this latter investigation.

Hong et al. (2017) use a canonical sender-receiver game with fixed roles and aligned interests. A random object is revealed to senders who then send a message to receivers. The receiver then guesses which object the sender saw. Both earn money if the receiver chooses correctly, and if not they do not. A main result of their paper is that perfect compositional languages emerges surprisingly regularly, from 50% to over 90% of the time, depending on the treatment. In contrast, Selten and Warglein (2007) report compositional languages to emerge only around 12% of the time, often less depending on treatment.

Hong et al. (2017) found convincing evidence that compositional language emerges. Further, on balance their data support the Blume (2005) hypothesis that compositional languages are learned more quickly than others. That said, Hong et al. (2017) were unable to assign language to participants randomly, but rather compared speed of learning between participants that learned different languages within the same treatment. The absence of randomization leaves open the possibility that their results are confounded by selection effects.

Our paper returns to the challenge to assess speed of learning across languages that differ in the number of observations required to identify them, and does so in a way that randomizes participants to languages. We use the Selten and Warglein (2007) paradigm, where previous studies suggest it is rather difficult for perfect languages to emerge. This works to our advantage as we can examine how different languages structures evolve and develop over time. Moreover, by randomly allocating participants to languages, we are able to rigorously test whether numbers of observations required to identify a language affects the speed with which a language can be learned.

3. A Model of Learning Language (Blume, 2005)

The goal of participants is to describe objects in a set Ω with messages from a set M, where each message is composed of symbols from a set S. We formalize this below. This formalization draws heavily from the development and notation of Blume (2005) and Hong et al. (2017).

The following features of the environment are common knowledge. The objects have a product structure: $\Omega = \times_{j=1}^{J} \Omega_j$, with $\#\Omega > 1$, and $M = \bigcup_{i=0}^{\infty} S^i$. M is therefore the union of cross-products of S. We set S to be a finite set of symbols where $S^i = S$ for all i $\in \{0,1,2,\dots\}$. We also require that $\#S \geq \max_j \#\Omega_j$. We define a language to be a one-to-one function $f\colon \omega \to M$. Then, we let $\mathcal{F}$ denote the set of all languages. An important regularity condition for speed of language results is that no language in $\mathcal{F}$ is focal.

We are interested how people learn languages from observation. We denote an observation from a language f by $(\omega, f(\omega))$. Thus, an observation is a two-vector whose elements are a state and the corresponding language-induced message. Observations reduce the set of possible languages. Let $\mathcal{F}(\omega, m)$ represent the group of all languages in $\mathcal{F}$ consistent with observation (ω, m). Following Blume (2005), learning from observations is defined as follows: A set of observations $\mathcal{O}$ induces the set of possible languages $\bigcap_{(\omega,m)\in\mathcal{O}} \mathcal{F}(\omega, m)$. We say a language is learned from observations $\mathcal{O}$ if the induced set of possible languages is a singleton.

Ease of Learning Hypothesis (Blume, 2005): A language f is said to be easier to learn than a language g if a (strictly) smaller number of distinct observations are needed to learn f.

It is useful to define compositionality. For a given language f, if messages can be decomposed into coordinate functions whose domains are the factors of the state space, i.e., $f(\omega) = \left(f_1(\omega_1), \dots, f_J(\omega_J)\right)$, where $f_j(\omega_j) \in S$ for all $j \in J$ and each f_j is an injection, then we say the language f is compositional.

Thus, if the meaning of any message in a language can be identified by combining the meanings of its symbols, then the language is compositional. Blume (2005) shows that as long as the objects have a product structure, there exists a compositional language that describes them and that requires the minimum number of observations to learn. Of course, not all compositional languages will achieve this lower bound.

For example, in the case of our experiment below, the objects to be described are emojis. The 12 emojis we consider can be characterized in two dimensions: (affect; arousal). Consequently, one type of compositional grammar would use a unique symbol for each affective state (in our case happy, sad and surprised), and one symbol for each arousal state (we include four levels, from low to high).

Alternatively, our emoji set can be organized into three dimensions (affect; arousal; red face). Under this structure, a compositional language would assign one symbol to each affect, one symbol to each of the (now three) levels of arousal, and one symbol to indicate the presence of a red face. While both grammars are compositional, we show below that it takes only three observations to identify the language in 2-dimensions, while it takes four to identify the language in 3-dimensions. Consequently, the framework above implies that the language in two-dimensions (which we incentivize in the 3-3 treatment below) should be more easily learned than the language in three-dimensions (incentivized in the 4-2 treatment).[7]

[7] Adding dimensions to the product space need not necessarily increase the number of observations required to learn a language, as this number depends both on features of the product space as well as number of available symbols. When the number of available symbols is sufficiently large, as it is in our case, learning can be more difficult than when the symbol space is restricted (Blume, 2005).

4. Experiment Design

We use emojis as objects and fruits as symbols. This differs from other laboratory experiments on emergent language (Blume et al, 1998; Selten and Warglien, 2007; Hong and Zhao, 2017; Hong et al, 2017). In those studies, the objects are geometric shapes and the symbols are either English letters or signs (e.g. !, @, #, $, %, ^). One reason we made this variation is to demonstrate that results on emergent compositional language are robust to types of objects and symbols used, since emojis are like most real-world objects, which not as obvious to be viewed as products of different dimension of features as geometric shapes. We did however want to preserve modularity of the objects in order to ensure that our experiment maps well to the motivating theory (Blume, 2005). The emojis we use have a simple product structure (e.g., affect, arousal) and thus serve well this purpose.

4.1 Treatments *3-3* and *4-2*

Our experiment includes two incentive conditions, which we denote by ***3-3** and **4-2***, both building from the experiment design introduced by Selten and Warglien (2007). Within each incentive structure people play a communication/coordination game, as described below.

Communication Game

The communication game involves two players. They both see the same list of n objects on their own screens, with order randomized and different for different subjects. For each object on the list, both players need to compose an expression using m symbols. Each player can choose any symbol(s) from the repertoire for each expression. The symbols may appear in any order and each may appear any number of times. It is not allowed to use the same expression to describe different objects or to leave a blank expression for any object.[8] When both players have submitted their expressions for all the objects, one object is randomly selected and the players' expressions for that object are compared. If the expressions perfectly match, the communication

[8] When a subject submits his/her expressions for all the objects in a round, if there is any identical expression used for different objects, or blank expression for any object, the program will display an error message to the subject and asks for a revision of the expressions until no violations of these rules are detected.

is successful and each of the subjects earns 10 EC, otherwise the communication fails and the subjects earn 0 EC.[9]

Regardless whether communication succeeds, exactly one randomly selected player must pay the cost for all symbols used in her own expression for the selected object. The other player does not pay any cost. Following the design of Selten and Warglien (2007), the randomization occurs at the end of each round. Therefore, when subjects make decisions regarding their expressions, both have 50% chance to have to pay costs, meaning that both should be sensitive to the cost of their fruit expressions.

At the end of each round, each subject receives the following information: the randomly selected object (emoji) and both subjects' expressions for that object within the pair, his/her randomly assigned role (sender/receiver), his/her current round's payoff, and the accumulated payoff from the first to the current round.

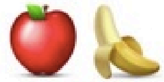
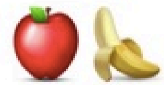
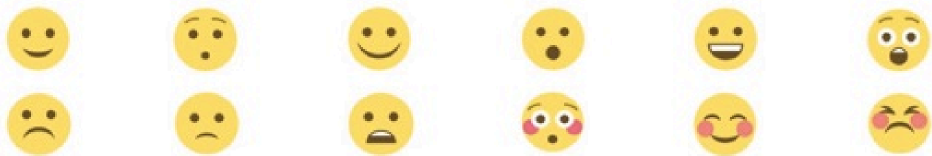
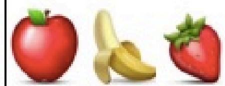
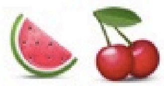
Treatments 3-3 and 4-2: Details

Participants play a 60-round communication game with emojis (e.g. 🙂, 😯) as objects, and fruits (e.g. 🍎, 🍇) as symbols. Our treatment conditions are denoted *3-3* and *4-2.* The full list of the objects (emojis) and the fruit repertoire with the costs, which vary between treatment, are listed in Table 1.

For the first 10 rounds, subjects only need to form expressions for two emojis using two fruits, each with cost 1 EC. From 11th round to 30th round, four more emojis are introduced into the object set to be described, while the size of the repertoire remains as before. From the 31st to the 60th round, six new emojis are added to the set, and we also make four additional fruits available for the expressions. Some of these new fruits cost 1 EC, and some cost 5 EC per use, as shown in Table 1. Beginning with few objects and symbols leaves it easier for subjects to create a language. With the two distinct cost structures in the two treatments, we expect different language structures to emerge in a precise way that we detail further below.

[9] Our game is a coordination game. The key difference between it and a standard coordination game is that the strategy space is infinite, in the sense that for each emoji the expression may in principle consist of an infinite number of fruits, which evidently leaves coordination more difficult *ex ante*.

Table 1. The emojis objects and the fruits repertoire (cost) in each round of the communication games in Treatment *3-3* and *4-2*

Round	Objects Treatment 3-3 & 4-2	Repertoire Treatment 3-3	 Treatment 4-2
1-10		 (1)	
11-30		 (1)	
31-60		 (1)  (5)	 (1)  (5)

Note: The number in the parentheses denote the cost per use of each fruit. E.g. ' (1)' means each use of  or  costs one EC.

The two incentive conditions are identical up to the cost of 🍇 which is 5 EC in treatment *3-3* and 1 EC in Treatment *4-2*. This means Treatment *3-3* includes three relatively lower-cost fruits (🍎,🍌,🍓) and three with higher costs (🍇 ,🍉 ,🍒) in the final 30 rounds, as compared to four lower-cost (🍎, 🍌,🍓 ,🍇) and two higher-cost fruits (🍉, 🍒) in Treatment *4-2*.

Subjects are endowed with 250 EC at the beginning of the 60-round communication games, and are informed that their initial 250 EC plus the ECs they earn during the communication games will be converted to US dollars at a rate of 1 EC= $0.03. Earnings were bounded below by zero (though as a practical matter this never occurred).

4.2 Language Predictions

In each part, we say a pair of subjects achieved a *common code* if they matched expressions for all 12 emojis in the final round of the communication game. A common code can be understood as an emergent language. Our predictions regarding the nature of the emergent language build from two assumptions. The first is that subjects prefer to use a language with lower average costs per emoji. Selten and Warglien (2007) report cost efficiency is an important factor for communication in their experiment.

Second, we expect, as found in the previous literature, emergent languages often to have a compositional structure. Tables 2(a) and 2(b) provide examples of compositional languages.

Participants are of course allowed to develop any coding structure. While experimental studies find simple (non-positional) compositional languages do emerge frequently (Selten and Warglien, 2007; Hong et al, 2017; Hong and Zhao 2017)[10], simple non-compositional frameworks are also observed (Hong et al, 2017). An example of a simple but non-compositional language to describe our 12 emojis is the repeat pattern detailed in 2(a), with the exception that the red-faced emojis are assigned their own repeating fruit pattern, such as one banana for the smiling red-face and two bananas for the sad red-face. Because such a repeat pattern is unrelated to the red-faced emojis product structure, the language is non-compositional.

[10] Another type of grammar, "positional compositionality", may further lower expression costs. We do not consider these as they are difficult for participants to discover, are not observed in our experiment, and are very rarely been observed in previous related experimental studies (Selten and Warglien, 2007; Kirby et al., 2008; Cornish et al., 2010; Hong et al., 2017).

Table 2. Examples of Compositional Language

(a) 2-Dimensional (2D) Compositional Language

Dimension 1 \ Dimension 2		Arousal Level			
		1st level single fruit	2nd level twice repetition	3rd level triple repetition	4th level four-time repetition
Affect	happy 🍎	🙂 🍎	☺️ 🍎🍎	😀 🍎🍎🍎	😊 🍎🍎🍎🍎
	sad 🍇	😕 🍇	☹️ 🍇🍇	😧 🍇🍇🍇	😣 🍇🍇🍇🍇

(b) 3-Dimensional (3D) Compositional Language

Dimension 1 \ Dimension 2(3)		Arousal Level/red face			
		1st level none	2nd level 🍌	3rd level 🍉	red face 🍒
Affect	happy 🍎	🙂 🍎	☺️ 🍎🍌	😀 🍎🍉	😊 🍎🍒
	sad 🍇	😕 🍇	☹️ 🍇🍌	😧 🍇🍉	😣 🍇🍒

Like all coordination games, anything on which people coordinate is an equilibrium. Based on the above two assumptions, however, it is easily verified that, given the cost structures, there is a unique cost-efficient simple (non-positional) compositional language in treatment 3-3, and two cost-efficient languages in 4-2, one compositional and one simple but non-compositional. These languages can of course be expressed in many different ways (by label switching, such as apples for bananas, for example). We predict the cost-efficient equilibria, examples of which are shown in Table 3, to emerge in our game.[11]

In treatment 3-3, where there are only three low-cost fruits avaialable, the most cost-efficient language uses one type of the cheap fruit only (🍎,🍌, 🍓), for each affect, and then repeating that fruit to indicate level of arousal. For instance, as shown in the example in Table 3, using different numbers of apples to describe a series emojis with a smiling face, using bananas for frowning faces, and strawberries for emojis with round-open mouths. Note that this languaage organizes the emojis according to two dimensions, affect and arousal, and thus we call this a 2D compositional language.

In order to identify the specific 2D compositional language at use one must observe which fruit is assigned to which emoji, and must also observe the rule for describing levels of arousal. This can be accomplished with three observations. For example, in treatment 3-3 described in Table 3 below, the language is identified by observing the emojis assigned to one apple, two bananas, and four strawberries. This determines the fruit assigned to each affect as well as the rule and ranking of levels of arousal (the omitted fourth category of arousal – the three repeat - can be inferred from the other observations). Note that this compositional language is learning efficient in the sense that there is no language, compositional or otherwise, that can be identified with fewer than three observations in this case (Blume, 2005).

For treatment 4-2, where there are four inexpensive fruits, the cheapest simple compositional language to express the 12 emojis uses a cheap fruit and its repetition for the emojis with same affect, but without the red faces. The red faced emojis are described by the combination of the fourth cheap fruit and the fruit associated with the emoji's affect. Note that this compositional

[11] This prediction is supported by previous experiments using similar games (Selten and Warglien, 2007; Hong et al., 2017; Hong and Zhao, 2017), and more generally experimental work with coordination games (e.g., Van Huyck et al., 1991) as well as theoretical work on coordination games including Crawford and Haller (1990); and Crawford and Sobel (1982).

Table 3. Predicted Cost-Efficient Emergent Languages in Treatment 3-3 and 4-2.

	Tretament 3-3	**Treatment 4-2**
	🍎🍌🍓 (1) 🍇🍉🍒 (5)	🍎🍌🍓🍇 (1) 🍉🍒 (5)
Predicted languages with lowest avg cost per emoji	2-D Compositional Language 🙂 🙂 😀 😊 — 🍎 | 🍎🍎 | 🍎🍎🍎 | 🍎🍎🍎🍎 🙁 ☹️ 😦 😣 — 🍌 | 🍌🍌 | 🍌🍌🍌 | 🍌🍌🍌🍌 😯 😮 😲 😳 — 🍓 | 🍓🍓 | 🍓🍓🍓 | 🍓🍓🍓🍓	Non-Compositional Language OR 3-D Compositional Language 🙂 🙂 😀 — 🍎 | 🍎🍎 | 🍎🍎🍎 🙁 ☹️ 😦 — 🍌 | 🍌🍌 | 🍌🍌🍌 😯 😮 😲 — 🍓 | 🍓🍓 | 🍓🍓🍓 😊 😣 😳 — 🍇 | 🍇🍇 | 🍇🍇🍇 OR 🍎🍇 | 🍌🍇 | 🍓🍇
avg cost per emoji	2.5	2

Note: The costs of the fruit sysmbols are displayed in the second row for each treatment.

structure organizes the language in three dimensions, (affect, arousal, red), and thus we refer to this language as 3D compositional. An equally cost-efficient simple but non-compositional language uses a repeat pattern with the inexpensive fruit for the red-faced emojis.

To identify that one is using the 3D compositional language one must know which fruit is assigned to each affect, as well as the rule for describing the red faced emojis. This requires four observations. For example, observing the emojis associated with one apple, two bananas, three strawberries along with the combination of apple and grapes identifies the language. To identify the simple non-compositional language requires five observations, as one must observe the message associated with two of the red emojis in order to know the third, because the "repeat grape" pattern does not take advantage of the emoji's product structure.

In the 3-3 treatmemnt learning efficiency and economic efficiency are aligned. Note however that the 4-2 treatment involves tension between learning efficiency and economic efficiency. The learning-efficient 2D language is of course available in the 4-2 treatment, but it is more economically costly than the less learning efficient 3D or non-compositional languages. We hypothesize that people are more sensitive to economic costs than learning efficiency, but our experiment allows us to observe how this tradeoff resolves.

4.3 Hypotheses

Hypothesis 1: The frequency with which language emerges will be the same in the 3-3 and 4-2 treatments.

This hypothesis asserts that changing the incentive structure of the environment will not change the frequency with which language emerges. Incentives should however impact the nature of emergent language, as we detail in Hypotheses 2 below.

Hypothesis 2a: An exact 2D compositional language emerges at least as frequently as the 3D compositional language in treatment 3-3.

Hypothesis 2b: An exact 3D compositional language emerges at least as frequently as a 2D compositional language in treatment 4-2.

Hypothesis 2c: An exact non-compositional language (repetition rule) emerges at least as frequently as a 2D compositional language in treatment 4-2.

Note that Hypotheses 2a-2c make one-sided predictions. Note further that these hypotheses are knife-edge, focusing on 2D or 3D languages that emerge perfectly. We can broaden these hypotheses as follows.

Hypothesis 3a: In treatment 3-3, the emergent languages will be "closer" to 2D compositional language than either a 3D or non-compositional (repetition rule) language.

Hypothesis 3b: In treatment 4-2, the emergent languages will be "closer" to a 3D compositional or repeat rule non-compositional than a 2D compositional language.

We explain the measurement of "closer" below. Note that these are also one-sided predictions.

Hypothesis 4: Learning is faster in 3-3.

4.4 Procedures

Experiments were conducted at the ICES lab of George Mason University in 2017. The subjects were recruited from the ICES subject pool consisting of undergraduate and graduate students from all backgrounds at George Mason University.

Subjects are randomly assigned to different treatments. A total of 72 and 64 individuals participated in our Treatments *4-2* and *3-3*, respectively. The communication games were implemented using software coded in Python, HTML, JavaScript and CSS.

Each subject's earnings include a $5 show-up fee at the beginning of the experiment and their performance-based earnings from the experiment. Each session on average takes 150 minutes for treatment 3-3 and 4-2. The average earnings are $36.5 in Treatment *3-3* and *4-4*.[12]

5. Results

To test Hypothesis 1, following Selten and Warglien (2007), we evaluate whether, for the 60 rounds of the communication game, each pair of subjects matched their expressions for all 12 emojis in the final round. Using the observations from the two parts of communication games including all 72 subjects (36 pairs) in treatment 3-3 and 64 subjects (32 pairs) in treatment 4-2, we obtain the following result:

Result 1. *Consistent with Hypothesis 1, the rate of language emergence does not differ across treatments.*

[12] In the experiment, after the 60-round communication game ended, we implemented it again (with some other tasks following). We randomly re-matched the subjects before the start of the second 60-round communication game. In the data analysis as will be shown, we only use the data from the first 60 rounds, for the obvious reason that the behavior in the second set of communication game may be influenced by subjects' experience from the first set.

We confirm this by calculating between-treatment differences in the fraction of subjects who have reached common code. In Treatment 3-3, 23 out of the 36 pairs (64%) of subjects achieved common code, and 15 out of the 32 pairs (47%) did it in Treatment 4-2. The between-treatment difference is insignificant (p-value=0.163 by a 2-sided t-test).

In order to test the remaining hypotheses, we focus exclusively on subjects who achieved a common code with their partners in the communication games.

Result 2. *Incentives influence emergent language in the way predicted by Hypotheses 2.*

In treatment 3-3, we observe 4 out of the 23 pairs (17%) that achieved common codes used exact 2D compositional languages, significantly more than the frequency (0%) of exact 3D compositional languages. The difference is significant (p=0.021 by a 1-sided t-test).[13] Hypothesis 2a is supported.

In treatment 4-2, we observe an exact 3D compositional language and an exact non-compositional language with a repetition rule (NC in the following), to emerge both with a frequency of 7% among all common codes, while an exact 2D compositional language never emerges. Although all the observed frequencies are small, the direction of the difference supports Hypothesis 2b and 2c.

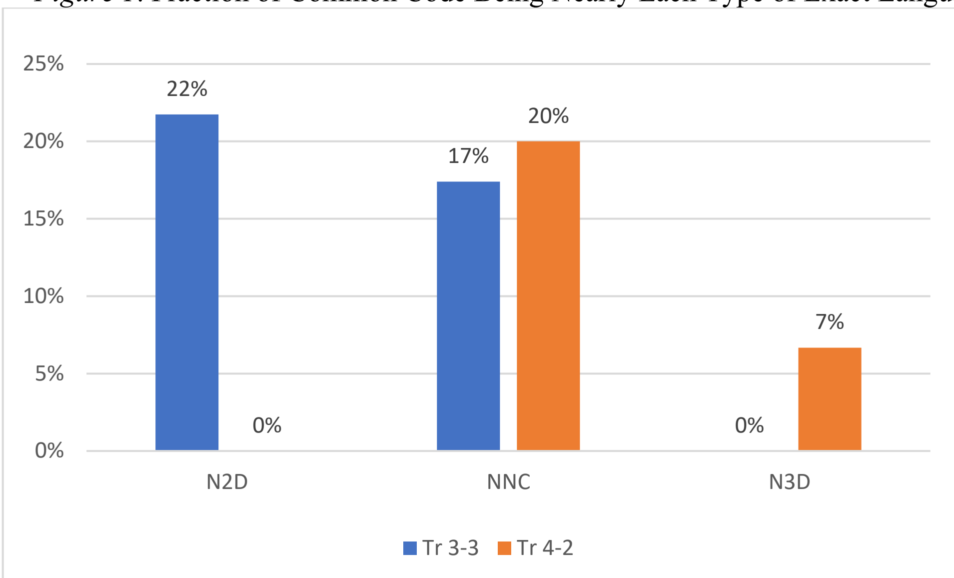
Result 3. *In treatment 3-3 more emergent languages are closer to 2D than 3D or NC languages. By contrast, in treatment 4-2 more emergent languages are closer to 3D or NC than 2D languages*.

This result is based on the following analysis. For all the common codes achieved by each pair of participants in our experiment, we calculate the distance between each pair's common code and each of the following three types of languages: exact 2D compositional language, exact 3D compositional language, and an exact non-compositional language (based on a repetition rule). We define the distance between a common code and an exact language to be the smallest number of changes required to the common code in order to create that exact language. Then, we say that common code is "nearly" one language (other than the two others) if the distance from that one exact language is smaller than the distances from the other two by at least 2. For example, if it requires two changes to make a particular common code into an exact 2D compositional language,

[13] The frequency of the 2D compositional language is also higher than that of the non-compositional (repetition rule) language (3 out of 23, or 14%), although the difference is not statistically significant (p =0.344 by 1-sided t-test).

but five for the other language alternatives, then we say that particular common code is a “nearly” 2D language (other than NC or 3D language).

Figure 1. Fraction of Common Code Being Nearly Each Type of Exact Language



The results of this analysis are summarized in Figure 1. We observe that in Treatment 3-3 and 4-2 a similar fraction of common codes are nearly non-compositional (NNC) language (17% and 20% respectively), but in Treatment 3-3, more than 20% of the common codes appear to be nearly 2D (N2D) language, while none are nearly 3D (N3D) language. In sharp contrast, in Treatment 4-2, no common codes are N2D, but 7% appear to be N3D.

This finding makes clear that the control of economic incentives we impose in different treatments effectively induces emergent languages with different structures.

Finally, we investigate whether common codes in 3-3 are closer on average to 2D languages, and 4-2 closer to 3D languages. To do this, we calculate for each common code the difference between its distance from the exact 2D language and its distance from the exact 3D language; we denote this variable as $\Delta_{dist(2D-3D)}$. Thus, a negative $\Delta_{dist(2D-3D)}$ indicates a common code closer to the 2D language than the 3D language, and vice versa. The mean of $\Delta_{dist(2D-3D)}$ is -0.52 in Treatment 3-3 and 0.27 in Treatment 4-2, and this difference is significant (p=0.044, 2-sided t test). This is further evidence that our treatment incentive design successfully influences the structure of the emergent languages, as hypothesized.

Result 4 (Support for Bloom, 2005). *Compositional languages that require relatively fewer examples to learn are learned more quickly than non-compositional or compositional languages that require relatively more examples to learn.*

Bloom (2005) predicts that it is faster to learn the 2D language than the 3D language. In light of the treatment effect on the emergent language structure indicated above, the implication of Bloom's theory is that the learning the language in Treatment 3-3 should be faster than in Treatment 4-2 (recall that the 2D language requires only three observations to learn, while the 3D requires four, and the non-compositional language requires five observations). We test this by comparing the speed of subjects in achieving language between the two treatments.

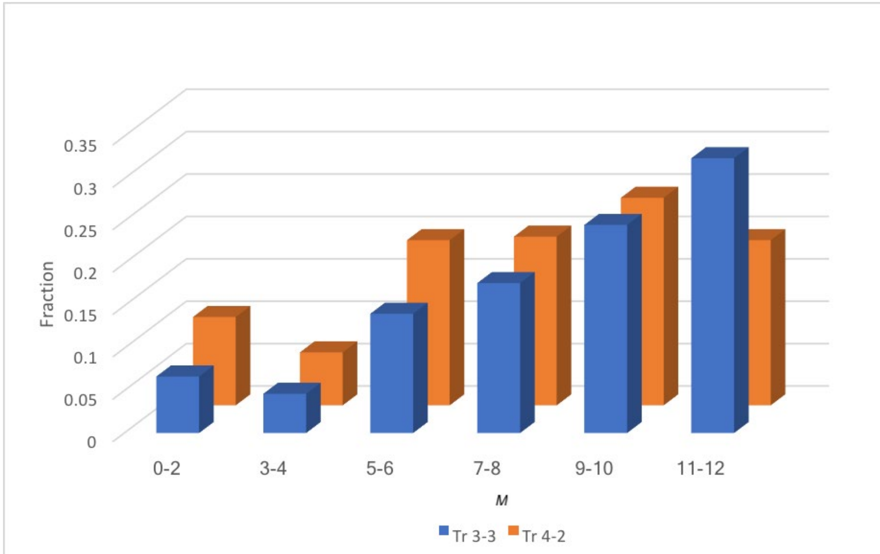
Our analysis proceeds as follows. We first determine for each treatment, the exact number of matches achieved by each pair i in round t of the communication game. We denote this by M_i^t. Then, 1) the distribution of M in one treatment serves as a natural measure of how successfully participants in a treatment learned a common code, with higher values indicating greater success; and 2) comparing the distribution of M between treatments reveals the difference in overall success of learning languages under the different incentive environments.

Recall that the two treatments' incentives (symbol costs) only differ in the last 30 rounds (Round 31-60) of the communication game. Consequently, we expect a difference in the distribution of M only in those rounds, and expect no significant difference in the previous rounds.

Figure 2 reveals, in support of Hypothesis 4, significant differences in the distribution of M between treatments during Round 31-60.[14] In particular, we observe that during the last 30 rounds of the communication game, subjects in Treatment 3-3 achieved more matches more often than in Treatment 4-2, and fewer matches less often in than in Treatment 4-2. The distributions across treatments are both visually and statistically significantly different (Kolmogorov-Smirnov test, $p<0.001$). In comparison, when the two treatments' incentives do not differ (during Round 11-30), we do not observe differences between the distributions of the number of matches between treatments ($p=0.365$ by K-S test).[15] Hypothesis 4 is supported.

[14] We coarsen the 13 possible outcomes {0, 1, …, 12} into 6 categories: 0-2, 3-4, 5-6, 7-8, 9-10, 11-12 in order to ensure a sufficiently large number of observations in each cell to perform reliable statistical inference.
[15] Round 1-10 are not considered for analysis.

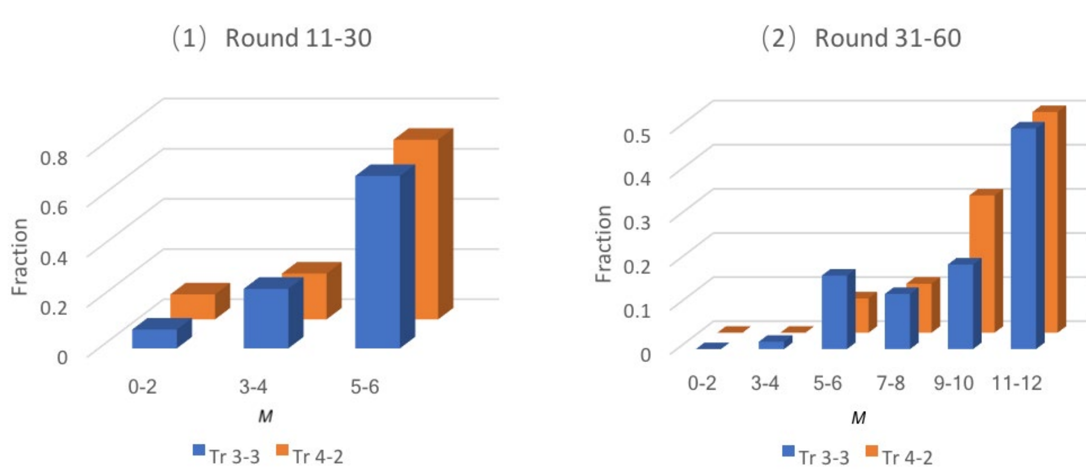
Figure 2. Distribution of the Number of Matches (M) during Round 31-60



Speed of Learning Non-Compositional Languages

Given the relatively large number of nearly non-compositional languages that emerged in both the 3-3 and 4-2 treatments, one may wonder whether between-treatment speed of learning differences in nearly non-compositional languages are driving our results. Figure 3 shows, however, that among groups that come up with nearly non-compositional languages, the distribution of M does not differ between treatments during either Rounds 11-30 or Rounds 31-60. A KS test confirms this (p=1.000, p=0.427, respectively). This is further support for Hypothesis 4.

Figure 3. Distribution of Number of Matches (M) among the Pairs Who Achieved NNC Languages



6. Conclusion

We developed a method to assign language randomly to participants in a laboratory experiment and used this approach to study the speed with which different languages can be learned. Our analysis was based on data from a communication game where participants coordinated on fruit expressions to describe emoji objects. Our paper offers both methodological and substantive contributions. Methodologically, we demonstrated that by varying the costs of symbols in the communication game one can reliably control the language that emerges from that game. This insight is valuable, as it opens the door to studies investigating links among, for example, language, culture, expectations and beliefs.

Substantively, our paper investigated whether, as hypothesized by Blume (2005), languages that can be identified with fewer observations can also be learned more quickly. This question has received recent attention (e.g., Hong et al, 2017) but conclusive findings remain elusive. Our ability to assign languages to participants randomly allowed for substantially sharpened results: we found clear evidence that a compositional language that requires three observations to identify is learned more quickly than compositional or non-compositional languages requiring four or five observations to identify. Finding support for Bloom's (2005) theoretical insight is important, as it contributes to explaining the universal use of compositional language in human society.

Acknowledgements

We are grateful to Jian Song for assistance with the experiments, to Hong Deng, Yuxuan He and Xin Wang for assistance with the data, and Kun Ma for assistance with programming. We received valuable feedback at early stages of this research from Xiaojian Zhao and Jie Zheng. For helpful advice and comments on earlier drafts we thank especially Joel Sobel and Wooyoung Lim, as well as workshop and seminar participants at Appalachian State University, Beijing Normal University, Central South University, Dongbei University of Finance and Economics, George Mason University, Harbin Institute of Technology, University of Tennessee, ShanghaiTech University, Shanghai University of Finance and Economics, and Tsinghua University. We are grateful to participants at the 2017 Camphor workshop in Dali, and the Brisbane 2018 ESA Asian meetings. Support for this research was provided by the Interdisciplinary Center for Economic Science at George Mason University, and research grants to the second author from the National Natural Science Foundation of China (#71703175, #71803201, #71773147) and the Guangdong Natural Science Foundation (#2017A030310392).

References

[1] Blume, A., DeJong, D., Kim, Y., & Sprinkle, G. (1998). "Experimental evidence on the evolution of meaning of messages in sender–receiver games". *American Economic Review*, 88(5), 1323–1340.

[2] Blume, A. (2000). "Coordination and learning with a partial language." *Journal of Economic Theory,* 95(1), 1-36.

[3] Blume, A., & Gneezy, U. (2000). An Experimental Investigation of Optimal Learning in Coordination Games. *Journal of Economic Theory*, 90(1), 161-172.

[4] Blume, A. (2005). "A learning-efficiency explanation of structure in language". *Theory and Decision*, *57*(3), 265-285.

[5] Bresnan, J. (1982). *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.

[6] Chan, C., Tardif, T., Chen, J., Pulverman, R., Zhu, L., & Meng, X. (2011). "English- and Chinese-learning infants map novel labels to objects and actions differently". *Developmental Psychology*, 47(5), 1459–1471.

[7] Cornish, H., Kirby, S., & Christiansen, M. H. (2010). "The emergence of structure from sequence memory constraints in cultural transmission." In *The Evolution Of Language* (pp. 387-388).

[8] Crawford, V., & Haller, H. (1990). "Learning how to cooperate: Optimal play in repeated coordination games." *Econometrica*, 58(3), 571—595.

[9] Crawford, V., Sobel, J., (1982). "Strategic information transmission". *Econometrica* 50, 1431–1451.

[10] Cremer, J., Garicano, L., & Prat, A. (2007). "Language and the theory of the firm". *Quarterly Journal of Economics*, 122, 373-407.

[11] Devetag, G. (2005). "Precedent transfer in coordination games: An experiment." *Economics Letters*, 89(2), 227—232.

[12] Franke, M., (2014). "Creative compositionality from reinforcement learning in signaling games." In: Cartmill, Erica A., et al. (Eds.), *The Evolution of Language: Proceedings of the 10th International Conference (Evolang 10)*. World Scientific, Singapore, pp. 82-89.

[13] Franke, M., (2016). "The evolution of compositionality in signaling games." *Journal of Logic, Language and Information*. 25, 355–377.

[14] Galantucci, B., Garrod, S., (2010). "Experimental semiotics: a new approach for studying the emergence and the evolution of human communication." *Interaction Studies* 11, 1–13.

[15] Galantucci, B., Garrod, S., (2011). "Experimental semiotics: a review. Front." *Human Neuroscience*. 17, 1–15.

[16] Hernández, P., Urbano, A., & Vila, J. E. (2012). Pragmatic languages with universal grammars. *Games and Economic Behavior*, 76(2), 738-752.

[17] Hong, F., Lim, W., and Zhao, X. (2017). "The emergence of compositional grammars in artificial

codes." *Games and Economic Behavior* 102:255-268.

[18] Hong, F., & Zhao, X. (2017). "The emergence of language differences in artificial codes." *Experimental Economics*, 20(4), 924-945.

[19] Kirby, S., Cornish, H., & Smith, K. (2008). "Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language." *Proceedings of the National Academy of Sciences*, 105(31), 10681-10686.

[20] Kyun, S. (2012). *Application of Contemporary Theories of Human Cognitive Architecture to the Design of Ill-defined Learning Domains: The Effect of Worked Examples when Learning English Literature.* (Doctoral dissertation, University of New South Wales Australia).

[21] Lipman, B. (2003). "Language and economics." In N. Dimitri, M. Basili, & I. Gilboa (Eds.), *Cognitive processes and rationality in economics*. London: Routledge.

[22] Marschak, J. (1965). "Economics of language." *Behavioral Science*, 10(2), 135–140.

[23] Rai, B. K., So, C. K., & Nicholas, A. (2012). "A primer on mathematical modelling in Economics." *Journal of Economic Surveys*, 26(4), 594-615.

[24] Rubinstein, A., (1996). "Why are certain properties of binary relations relatively more common in natural language?" *Econometrica* 64, 343–356.

[25] Rubinstein, A., (2000). *Economics and Language*. Cambridge University Press, Cambridge.

[26] Selten, R. and Warglien, M., 2007. "The emergence of simple languages in an experimental coordination game." Proceedings of the National Academy of Sciences 104(18):7361-7366.

[27] Steinert-Threlkeld, S., (2016). "Compositional signaling in a complex world." *J. Logic, Lang. Inf*. 25 (3), 379–397. December.

[28] Suzuki, T. (2020). "Efficient communication and indexicality." *Mathematical Social Sciences*.

[29] Szabó, Z.G., (2013). "Compositionality." In: Zalta, Edward N. (Ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2013 edition.

[30] Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1991). "Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games." *Quarterly Journal of Economics*, 106, 885–911.

[31] Weber, R., & Camerer, C. (2003). "Cultural conflict and merger failure: An experimental approach." *Management Science*, 49, 400–415.