The Emergence of Compositional Grammars in Artificial Codes *

Fuhai Hong[†] Wooyoung Lim[‡] Xiaojian Zhao[§]

November 22, 2016

Abstract

This paper experimentally explores how compositional grammars in artificial codes emerge and are sustained. In a communication game with no conflict of interest, the sender sends a message that is an arbitrary string from available symbols with no prior meaning to indicate an abstract geometrical figure to the receiver. We find strong evidence from the laboratory for the emergence of compositional grammars in the subjects' common codes that facilitate learning efficiency. Moreover, when there is a scarcity of symbols in the repertoire, a few groups in our experiments developed languages with positional compositionality, meaning the same symbol has different interpretations depending on its position in a string, whereas some other groups developed language structures that are not compositional but still efficient in communication.

Keywords: Communication Games, Economics of Language, Experimental Semiotics, Compositional Grammar.

JEL classification: C72, C92, D03, D83

*We are deeply indebted to Andreas Blume for his constructive comments that improved the paper immeasurably. We would also like to thank Vincent Crawford, the advisory editor, three anonymous referees, Te Bao, Andreas Ortmann, Joseph Tao-yi Wang, Eric van Damme, Yang Yang, and Songfa Zhong, the conference participants at the 2nd Behavioral and Experimental Economics Workshop (Chengdu), the Annual Congress of the European Economic Association (Toulouse), the Asian Meeting of the Econometric Society (Taipei), the ESA World Meeting (Sydney), the Fudan University Workshop on Behavioral Science (Shanghai), the International Conference on Economics and Econometrics (Hiroshima), and seminar participants at the City University of Hong Kong, Hong Kong University of Science and Technology, Nanyang Technological University, Shandong University, Shanghai University of Finance and Economics, Singapore Management University, Southwestern University of Finance and Economics, Tsinghua University and Wuhan University for their helpful comments. This study is supported by a grant from the Research Grants Council of Hong Kong (Grant No. ECS-699613).

[†]Division of Economics, Nanyang Technological University, fhhong@ntu.edu.sg.

[‡]Department of Economics, Hong Kong University of Science and Technology, wooyoung@ust.hk.

[§]Department of Economics, Hong Kong University of Science and Technology, xjzhao81@gmail.com.

1 Introduction

The key component of human language that enables us to efficiently transmit potentially complex information is its grammar. Despite the diversity in natural languages, economic theorists typically argue that there are common properties that guide the emergence and development of languages. Rubinstein (1996) is among the first to use the economic approach to derive linguistic properties. Rubinstein (2000) asserts that language is a mechanism of communication, and because economic theory analyzes the design of social systems, it must be relevant to linguistic issues as well. As tools of communication, languages arguably possess certain common properties that maximize efficiency in communication.[1] Although the theoretical advancement in this line of research has been well documented, the empirical test of the claimed commonality of linguistic properties is still in its infancy.

In this paper, we explore, on an experimental basis, how the grammar-like structure of language emerges and present evidence from the laboratory. We focus on the emergence of a linguistic property called *compositionality* (Frege, 1882), which appears to be universal in natural languages. The compositionality of languages suggests that, as strings of symbols, messages should be decomposable into coordinate functions whose domains are the factors of the state space.[2] The consideration of compositionality is associated with the notion of learning efficiency, which implies that available symbols are used such that individuals are able to learn the language efficiently, i.e., from a minimal number of observations. Blume (2005) shows that compositional languages are efficient in learning.[3]

We study how compositionality arises to facilitate communication efficiency in the following simple environment. Two players, a sender and a receiver, play a communication game with no conflict of interest. The sender, who is privately informed about the underlying state of nature, sends a message to the uninformed receiver, who then identifies the realized state. If, and only if, the receiver correctly identifies the state, both receive a positive payoff. The messages are strings of symbols with arbitrary lengths, where the available symbols have no a priori meaning.

In our laboratory implementation, we considered a state space with a $3 \times 2 \times 3$ product structure. We used abstract geometrical figures to denote the states of nature and allowed individuals to communicate using abstract symbols with no prior meaning to minimize the impact of any existing communication convention among the participants. Specifically, we used shapes such as ◯, △ and □ to denote objects and used vertical and horizontal orderings

[1] See also the follow-up survey by Lipman (2003).

[2] This is the definition of compositionality considered in the economics literature (refer to, e.g., Blume, 2005; Selten and Warglien, 2007). We discuss broader definitions of compositionality available in the linguistics literature in Footnote 12 and their implications for our experimental results in Footnote 27.

[3] Blume (2000) shows that grammatical structures facilitate coordination and learning in an environment in which novel meanings should be expressed. In philosophy and psychological linguistics, Franke (2014, 2016) and Steinert-Threlkeld (2016) consider the simple sender-receiver games proposed by Lewis (1969) to theoretically show that the compositionality of languages can emerge for unsophisticated agents from a variant of reinforcement learning.

of the shapes to denote the spatial relations between objects.[4] The number of available symbols in a message string was our main treatment variable. In our first set of treatments, senders were allowed to send a message consisting of a string with arbitrary length from five available symbols !, @, #, $ and %. In our second set of treatments, there were three available symbols !, $ and % only.[5]

With the above symbols, however, there may be focal associations between certain symbols and certain shapes, e.g., @ may be focal to be used to represent ◯. These focal associations may affect the speed of learning. To address this issue, we designed robustness treatments in which the sender's message was privately translated into a message with English letters, following some randomly determined rule, before being sent to the receiver. This *privatized translation of message* protocol is initially developed by Blume, DeJong, Kim and Sprinkle (1998) to remove all conceivable language conventions that may be shared among players.

We obtained strong evidence for the emergence of compositional grammars. Most groups in our first set of treatments (with five symbols) and more than half of the groups in our second set of treatments (with three symbols) developed compositional languages. We also observed that some symbols were never used in some groups because compositionality saved the use of additional symbols even if using them is costless. Interestingly, three groups in the second set of treatments developed languages with the property of "positional" compositionality, meaning that the same symbol has different interpretations depending on its position in a string. More importantly, compositionality indeed facilitated learning efficiency. Most groups learned the entire language based on a few observations in the beginning rounds. That is, participants were indeed able to rely on the structure of language to successfully communicate states they have not previously communicated. Moreover, the groups that did not develop compositional languages in the second set of treatments developed certain language structures as well that also facilitate communication. We found that most of our qualitative findings are robust to the introduction of the privatized translation of message.

The paper is organized as follows. The rest of this section reviews the related literature on language and economics. In Section 2, we discuss preliminaries that define the compositionality of languages and link it to the notion of learning efficiency. Section 3 introduces the experimental design and procedure. Experimental results are reported in Section 4. Concluding remarks are presented in the last section.

[4]In our experiment, we focused only on spatial relations, although in principle, relations can be more broadly demonstrated. Similarly, Kirby, Cornish and Smith (2008) introduce a variety of spatial relations and motions and study the origins of structure in human language. To study the non-intentional evolution of language structure, the authors focus on transmission errors in an iterated learning process via a diffusion chain, a standard methodology for studying information and culture transmission in experimental psychology.

[5]Due to technical difficulties, we were only able to implement arbitrary strings from the available symbols with any length of up to 20 characters in the laboratory. However, we argue that a length of up to 20 characters is sufficiently long because the number of states in our treatments was 18.

1.1 Related Literature

There is a small but growing literature in experimental economics on the emergence and evolution of language. Among others, the studies by Blume et al. (1998) and Selten and Warglien (2007) are most closely related to ours.[6]

Blume et al. (1998) experimentally explore the endogenous emergence of meanings in various sender-receiver games with two types considered by Lewis (1969). The authors are pioneers in providing a fully controlled laboratory environment in which there is complete absence of a shared language at the outset. Following Blume et al. (1998), we considered the case with no shared languages and studied the emergence of meaning in communication. However, our main focus was on how language *structure* emerges and whether the emerged language structure satisfies certain properties.[7]

Selten and Warglien (2007) present experimental evidence of the emergence of compositional languages in the laboratory. There are a number of similarities between the present study and that of Selten and Warglien (2007). In both studies, the research question is whether common codes can be developed and whether any structure of language such as compositional grammar emerges, and the research focus is on the structural properties of artificial languages but not on the nature of signs. Moreover, experimental subjects begin with no common language and must develop an artificial common code in the course of interaction to communicate about the state of nature, which comprises geometrical figures on a computer screen. Finally, both studies investigate the treatment effect of the size of the repertoire of permissible symbols, where the set of permissible symbols is limited.

Our experimental design is fundamentally different from that of Selten and Warglien (2007) in three respects. First, Selten and Warglien (2007) assume positive but finite costs of using an additional letter in a message and study how the costs and benefits of communication affect the emergence of the compositional grammar. In our experiment, the length of a message is arbitrary, and the use of available symbols is costless, as in natural languages. Second, the experiments of Selten and Warglien (2007) begin with a small number of states and a small number of available letters, and the set of states and the available (but costly) letters change over time. In contrast, our set of states and set of available symbols are fixed throughout the experimental session. Third, players' roles (either sender or receiver) are fixed in our experiments, whereas they are random in each round in the study of Selten and Warglien (2007).[8]

The highlighted differences in experimental design between ours and that of Selten and

[6]See also Weber and Camerer (2003) on the emergence of a common code based on natural language in the context of organizational culture and mergers.

[7]In experimental philosophy, Bruner, O'Connor, Rubin, and Huttegger (2014) also present experimental evidence for the emergence of meaning in the framework of the sender-receiver games considered by Lewis (1969).

[8]There is one more difference in the experimental design. In the work of Selten and Warglien (2007), senders adopt strategies prior to observing the state. We follow Blume et al. (1998) by using a choice method in which the sender sends a message in response to the privately observed state.

Warglien (2007) may translate to differences in experimental results.

First, in the study by Selten and Warglien (2007), with a small repertoire of symbols, subjects did not develop a common language, although in theory subjects can potentially rely on positional compositionality to facilitate their communication. In contrast, the small number of available symbols in some of our treatments was not an obstacle to efficient communication. We observed that three groups in these treatments developed languages with positional compositionality. Most other groups developed various types of common codes with efficient communication. The potential difficulty of developing a common code due to the lack of sufficient available symbols was mitigated by clever uses of the symbols, such as repeating the same symbol many times, because increasing the length of the message string was costless for the subjects in our study.[9]

Second, in the main experiments of Selten and Warglien (2007), only 12% of the groups developed compositional languages. By contrast, we observed the emergence of compositional grammars from a majority of groups in our treatments. One possible reason is that Selten and Warglien (2007) divided their experiment into three stages and the complexity of the state space and the set of available letters varied in the later stages. With a simple state space in the early stages, a non-compositional language could be nearly as efficient in learning as a compositional language. As the state space gradually becomes more complex, there is a marginal benefit of making the language grammatical, but there could also be a cost associated with changing the structure of language because of the "lock-in" effect; it may not be easy to introduce a grammar into the original non-grammatical language. However, in our design the state space was complex and fixed from the beginning, causing subjects to find a compositional language appealing. The subjects had a greater incentive to develop a grammatical language once and for all.

Our methodology is borrowed from the emerging field of experimental semiotics, particularly semiotic matching games. Semiotics is, in essence, the science of symbols and signs. Experimental semiotics focuses on interactions in the absence of any pre-established communication conventions and establishes the origins of novel languages. The experimenter can manipulate and control the subjects' environment, which would be difficult to do in the field. In semiotic matching games, both the set of referents and the set of forms that subjects use for communication are closed. With these closed sets, it is more feasible to investigate language-like structure or grammar in the laboratory. The literature has covered the issue of combinatoriality (Galantucci et al, 2010) and compositionality (see Selten and Warglien, 2007 as discussed above) of language. For a review of this literature, see, e.g., Galantucci and Garrod (2010, 2011).

[9]Moreover, we incorporated fixed and asymmetric roles of players that further facilitated the subjects to develop their common codes.

2 Preliminaries

We study a canonical communication game without conflicts of interests between players. There are two players, a sender and a receiver. The receiver is uncertain about the state of nature ω in a finite set Ω. She knows the prior distribution of states, F, with probability p_ω for $\omega \in \Omega$. The sender privately observes the realized state of nature and sends a message m to the receiver from a set of feasible messages M. After receiving the message, the receiver is asked to indicate which state is realized. If the receiver indicates the state correctly, communication is successful, and both players obtain a payoff normalized to one; otherwise, both receive nothing. Both players have von Neumann-Morgenstern utility.

The environment we consider in the present paper relies heavily on that of Blume (2005).[10] Assume that $\Omega = \times_{j=1}^{J}\Omega_j$ with $\#\Omega > 1$, and $M = \bigcup_{i=0}^{\infty} S^i$, i.e., M is the union of cross-products of S, where S is a finite set of symbols with $S^i = S$ for all i, and $\#S \geq \max_j \#\Omega_j$. Following Blume (2005), we define a language as a one-to-one function $f : \Omega \rightarrow M$. Let $\mathscr{F}$ denote the set of possible languages. $\mathscr{F}$ is the shared knowledge of players on languages they consider possible. We further assume that there is no language in $\mathscr{F}$ that is more focal than others. This assumption requires that all available symbols are *a priori* symmetric and therefore no association of symbols with components of the state space can be ruled out *a priori*. The complete symmetry of the symbols necessitates learning because the assumption implies that for any language in the set of possible languages $\mathscr{F}$, any other language that can be obtained via a permutation of symbols also has to be in $\mathscr{F}$. All of these factors are common knowledge.

To formally define learning efficiency as a motivation for the compositionality of languages, define an observation from a language f, denoted by $(\omega, f(\omega))$, as a pair consisting of a state and the corresponding message induced by the language. Observations break symmetries and thereby reduce the set of remaining languages that can be considered possible.[11] Let $\mathscr{F}(\omega, m)$ denote the set of all languages in $\mathscr{F}$ that are consistent with the observation (ω, m). Blume (2005) defines learning from observations as follows: A set of observations $\mathcal{O}$ induces the set of possible languages $\cap_{(\omega,m)\in\mathcal{O}}\mathscr{F}(\omega, m)$. A language can be learned from $\mathscr{F}$ via a set of observations $\mathcal{O}$ if the induced set of possible languages is a singleton. 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. A language f is said to be *efficient in learning* if there is no other language in $\mathscr{F}$ that is easier to learn than f.

We now define the *compositionality* of languages as follows:

[10]Blume (2000, page 9) also considers a similar setup in which a sender and a receiver repeatedly interact. In each interaction, the privately informed sender sends a message as a string of symbols to the receiver while there is no *a priori* focal language shared between the two players. The type space in Blume (2000, page 9) is a strict subset of a product space, which makes it possible always to learn the language with one observation.

[11]Crawford and Haller (1990) is the first to investigate situations with absence of a common language, which is modeled via symmetry constraints. They apply symmetry constraints to learning in repeated coordination games and show that players use precedents to achieve and maintain coordination.

Definition 1 (Compositionality) *A language f is **compositional** if messages can be decomposed into coordinate functions whose domains are the factors of the state space, i.e., $f(\omega) = (f_1(\omega_1), ..., f_J(\omega_J))$, where $f_j(\omega_j) \in S$ for all $j \in J$ and each f_j is an injection.*

When the state space has a product structure with two dimensions and the messages are strings of two symbols, for instance, one symbol in the message string describes one dimension of the state space, and the other symbol describes the other dimension independently in a compositional language. The idea is that the meaning of a message is a combination of the meanings of its symbols (Frege, 1882).[12]

Blume (2005) shows that if the state space has a product structure and messages are strings of symbols, a language with a compositional structure is efficient in learning. Put differently, the compositionality of a language qualifies the language as being learning efficient.[13]

Proposition 1 (Blume, 2005) *When $\Omega = \times_{j=1}^{J} \Omega_j$ with $\#\Omega > 1$ and $M = \bigcup_{i=0}^{\infty} S^i$ where $S^i = S$ for all i, S is a finite set of symbols and $\#S \geq \max_j \#\Omega_j$, a compositional language $f : \Omega \to M$ is efficient in learning.*

3 Experimental Design and Hypotheses

For our experimental implementation, we consider the following simple environment. Let $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$, where $\Omega_1 = \Omega_3 = \{a, b, c\}$ and $\Omega_2 = \{\alpha, \beta\}$. Our experimental design consists of two baseline treatments – treatments $U5$ (**U**nlimited-**5** symbols) and $U3$ (**U**nlimited-**3** symbols) – and two corresponding robustness treatments – treatments $U5$-PTM and $U3$-PTM, where PTM refers to the *privatized translation of messages*, an additional protocol we implemented in the robustness treatments which will be further elaborated later in this section.[14]

In the treatments $U5$ and $U5$-PTM, $M = \bigcup_{i=1}^{20} S^i$, where $S^i = S$ for all i and $\#S = 5$. More precisely, a message is an arbitrary string from five symbols – !, @, #, \$, and % – with any

[12] Note that this is one version of compositionality. There are alternative, broader definitions of compositionality available in the linguistics literature. For example, in the work of Franke (2016), a language is said to be compositional if the meanings of its composite expressions are systematically derived from the meanings of their parts and the way in which these parts are combined. In the study by Steinert-Threlkeld (2016), a language is said to be compositional if the meaning of a complex expression is determined by the meaning of its parts and how the parts are put together. For more details, see Janssen (2016) and Szabó (2013). These alternative definitions have certain implications for our experimental results, which will be discussed in Section 4.

[13] Blume (2005) also considers learning from fragments as well and shows that compositionality is ultimately both necessary and sufficient for learning efficiency.

[14] In the previous version of the paper, we also considered two treatments in which the message is *limited* in length and on which symbols could be used in a given string position. The results pertaining to the emergence of compositional grammars are robust in the treatments with limited messages, which is reported in Hong, Lim and Zhao (2015) and available upon request. Our naming choice for the treatments we report in the present paper was established to highlight the fact that the message space is unlimited in the above two respects for these treatments.

length between 1 and 20 characters. In the treatments $U3$ and $U3$-*PTM*, $M = \bigcup_{i=1}^{20} S^i$, where $S^i = S$ for all i and $\#S = 3$. A message is an arbitrary string from three symbols – !, \$, and % – with any length between 1 and 20 characters as well. The state space is $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$, and each of the 18 states ω has equal probability $p_\omega = \frac{1}{18}$ to occur. The length limit of 20 characters imposed on the message space is due to technical difficulties encountered in z-Tree in implementing the unlimited length. Nonetheless, this length is sufficiently long because $\#\Omega = 18$.

In the baseline treatments, the message string chosen by a sender was *directly* transmitted to the receiver in the same group. In the robustness treatments, however, we completely destroyed all conceivable focal associations between symbols and components of the state space that players might use for a priori coordination by implementing the *privatized translation of messages* (*PTM*), the protocol initially developed by Blume et al. (1998). Precisely, after a sender input his/her message, each distinct symbol in the written message was translated into one distinct letter among D, E, F, G and H in the treatment $U5$-*PTM*. For each group, the symbol-letter translation mapping was drawn uniformly randomly from the set of all bijections from S into $\{D, E, F, G, H\}$ at the beginning of the experiment and was not revealed to the subjects. But, once uniquely determined, the same symbol-letter translation mapping was used throughout the entire session for each group. An analogous procedure was used for the treatment $U3$-*PTM*. For more details, see the experimental instructions in Appendix A.

There is one important remark regarding possible focal associations between symbols and components of the state space in the treatments $U5$ and $U3$. For example, one may naturally argue that in the treatment $U5$ using symbol @ to represent ◯ and # to represent □ looks more focal because of the similarity in shapes. Even if the two symbols @ and # were made unavailable in the treatment $U3$, other possible focal associations may still exist unless the *PTM* procedure is adopted. Note that having focal associations between symbols and components of the state space is akin to having observations, and thus, reduces $\mathscr{F}$. Hence, all else being equal, having (potential) focal associations as in the treatments $U5$ and $U3$ should weakly speed up learning, relative to the case in which there are no focal associations as in the treatments $U5$-*PTM* and $U3$-*PTM*.[15]

3.1 Predictions and Hypotheses

Consider the language illustrated in Table 1 in the environment described in our treatment $U5$ ($U5$-*PTM*). In this table, a language is described by a panel of three matrices, where the choice of matrix refers to the element of Ω_1, the choice of row refers to the element of Ω_2 and the choice of column refers to the element of Ω_3. For example, the top-left cell in the middle

[15]More precisely, having some (but not full) focal associations between symbols and components of the state space is akin to having "partial" observations rather than full observations, which is associated with learning from fragments. As Blume (2005) shows, when learning from fragments is possible, the compositionality becomes a necessary condition for the efficient learning.

$a \in \Omega_1$

	a	b	c
α	!@!	!@$	!@%
β	!#!	!#$	!#%

$b \in \Omega_1$

	a	b	c
α	$@!	$@$	$@%
β	$#!	$#$	$#%

$c \in \Omega_1$

	a	b	c
α	%@!	%@$	%@%
β	%#!	%#$	%#%

Table 1: A Compositional Language in Treatment $U5$ ($U5$-PTM)

matrix presents the message "$@!" used to represent the state $b\alpha a$. Clearly, the language has a compositional grammar because there is a one-to-one mapping from the set of available symbols to the set of factors of the state space as follows:

$$! \to a, \$ \to b, \% \to c, @ \to \alpha, \text{and } \# \to \beta, \tag{1}$$

and the meaning of a message is composed of the meanings of the symbols in the message. We formulate the first hypothesis as follows:

Hypothesis 1 (Compositionality) *Compositional languages emerge in the treatments* $U5$ *(*$U5$*-PTM) and* $U3$ *(*$U3$*-PTM).*

$a \in \Omega_1$

	a	b	c
α	!$!	!$$	!$%
β	!%!	!%$	!%%

$b \in \Omega_1$

	a	b	c
α	$$!	$$$	$$%
β	$%!	$%$	$%%

$c \in \Omega_1$

	a	b	c
α	%$!	%$$	%$%
β	%%!	%%$	%%%

Table 2: A Compositional Language in Treatment $U3$ ($U3$-PTM)

An important difference in the treatment $U3$ ($U3$-PTM) from the treatment $U5$ ($U5$-PTM) is that $\#S = 3 < \#\Omega_1 + \#\Omega_2 = 5$, i.e., the sum of the number of elements in Ω_1 and the number of elements in Ω_2 that need to be labeled in a compositional code is strictly larger than the number of available symbols in the treatment $U3$ ($U3$-PTM). Thus, it is impossible to have a similar one-to-one mapping from the set of available symbols to the set of factors of the state space as in (1) under the treatment $U3$ ($U3$-PTM). However, it is still possible to have a compositional grammar as described in Table 2. The mapping presented here is a mapping from the Cartesian product between the *positions* of symbols in a message string and the set of available symbols to the set of factors of the state space:

$$!\_\_ \to a, \$\_\_ \to b, \%\_\_ \to c,\ \_\_! \to a,\ \_\_\$ \to b,\ \_\_\% \to c, \_\$\_ \to \alpha,\ \text{and } \_\%\_ \to \beta,$$

where, for example, !__, _!_ and __! represent the symbol ! positioned in the first, second, and third place, respectively, of a message string with a length of three.

This type of compositionality is called positional compositionality, meaning that the same sign has different interpretations depending on its position in a string (see, e.g., Selten and Warglien, 2007). Clearly, the mapping required by such positionally compositional grammars is more complicated than the standard "*simple*" compositional grammars.[16] Thus, it is natural to imagine that subjects would have a difficult time utilizing it. Moreover, the limited number of available symbols in the treatment $U3$ ($U3$-PTM) creates a higher demand for the positional compositionality. Taken together, we formulate our second hypothesis as follows:

Hypothesis 2 (Positional Compositionality) *Compositional grammars emerge more often in the treatment U5 (U5-PTM) than in the treatment U3 (U3-PTM). Among the groups developing compositional grammars, positionally compositional grammars emerge more often in the treatment U3 (U3-PTM) than in the treatment U5 (U5-PTM).*

Regarding the learning efficiency of compositional languages, as demonstrated by Blume (2005), the languages presented in Tables 1 and 2 are efficient in learning. Taking the language in Table 1 for example, if the individual knows the compositional structure of the language but does not know exactly which symbol represents which factor, at most three distinct observations are sufficient for him to learn the entire language.[17] We thus formulate the following hypothesis.

Hypothesis 3 (Learning Efficiency of Compositional Languages) *Groups developing compositional languages learn the languages with at most three distinct observations.*

Our next hypothesis concerns about the efficiency gain from more permissible symbols. If compositional languages are efficient in learning (Hypothesis 3) and such languages emerge more frequently in the treatment $U5$ ($U5$-PTM) than in the treatment $U3$ ($U3$-PTM) (Hypothesis 2), then one may expect that the efficiency gain from more frequently emerging compositional languages in the treatment $U5$ ($U5$-PTM) is translated into higher success rate of communication.

Hypothesis 4 (Efficiency Gain from More Permissible Symbols) *The average frequency of successful communication in the treatment U5 (U5-PTM) is higher than that in the treatment U3 (U3-PTM).*

Our last hypothesis concerns about possible focal associations between symbols and components of the state space in the treatments $U5$ and $U3$. Note that the privatized translation protocol is adopted to get rid of any existing language convention and to provide the initial environment with perfect symmetry. As discussed in the previous section, the absence of perfect symmetry accelerates learning.

[16] We thus say a language satisfies the *simple* compositionality if it satisfies the standard compositionality introduced in our Definition 1 but not the positional compositionality. We thank Andreas Blume who proposed this term to us.

[17] However, three different observations are not always necessary. For example, two proper observations, e.g., $(a\alpha a, !@!)$ and $(a\alpha b, !@\$)$, are sufficient.

Hypothesis 5 (Effect of Privatized Translation of Messages) *Groups in the treatment $U5$ ($U3$) learn the languages faster than those in the treatment $U5$-PTM ($U3$-PTM).*

3.2 Experimental Procedures

This subsection describes the experimental procedures. More details can be found in the instructions in Appendix A. For each treatment, two sessions were conducted in English in the experimental laboratory at the Hong Kong University of Science and Technology (HKUST) in May and August, 2016. Each session for the baseline treatments featured 20 participants while the numbers of participants for the robustness treatments varied between 8 and 14. We used a between-subject design. In total, 122 students, none of whom had any prior experience with our experiment, were recruited as subjects.[18] All sessions were conducted using z-Tree (Fischbacher, 2007).

To visualize the states of nature to the participants, we used abstract geometrical figures to denote the states. We used △ to denote a, ◯ to denote b, and □ to denote c; we let the horizontal ordering between two objects denote α and let their vertical ordering denote β. The state space contained 18 equally likely figures. Table 3 summarizes our design of the treatments.[19]

Treatment	States Space									Message Space	*PTM*
$U5$	△	△	△	◯	◯	◯	□	□	□	$M = \bigcup_{i=1}^{20} S^i$	No
$U5$-*PTM*	△	◯	□	△	◯	□	△	◯	□	where $S = \{!, @, \#, \$, \%\}$.	Yes
$U3$	△△	△◯	△□	◯△	◯◯	◯□	□△	□◯	□□	$M = \bigcup_{i=1}^{20} S^i$	No
$U3$-*PTM*										where $S = \{!, \$, \%\}$.	Yes

Note: *PTM* refers to the *privatized translation of messages.*

Table 3: Experimental Design

At the beginning of each session, roles were randomly assigned, and one sender and one receiver were randomly paired. Subjects' roles as well as their paired partners were fixed throughout the experiment so that every data point is from an interaction between two individuals.[20] Subjects in each group interacted for 50 rounds.[21] At the beginning of each round, both subjects in each group were shown the set of figures on the computer screen. The sender

[18]The number of participants for each treatment was 40, 20, 40, 22 for treatments $U5$, $U5$-*PTM*, $U3$, and $U3$-*PTM*, respectively.

[19]In our experiments, the relative positions of the figures were displayed in a random manner both in the instructions and on the computer screen for each player in each round during play.

[20]The fixed matching protocol was used to facilitate the emergence of efficient communication in a limited number of interactions in the lab. The fixed role design has been widely used in the literature to study the emergence of meaning in sender-receiver games (e.g., Blume et al., 1998).

[21]This large number of rounds is sufficient for us to observe the converged behaviors. In all treatments, the receiver observed the chosen figure at the end of each round. Our repeated-interaction setting is particularly suited to studying the learning efficiency of language in that the receiver learns by trial and error, at least for the beginning periods of our game. Therefore, economic incentives are provided for individuals to develop a learning-efficient language to minimize learning costs.

was then shown the chosen figure and asked to send a message to the receiver in the same group. The message from the sender was transmitted directly to the receiver in the baseline treatments, or translated via the privatized translation of message procedure described above and transmitted to the receiver in the robustness treatments. Upon receiving the message, the receiver was asked to indicate which figure was chosen. After the receiver's decision was made, feedback on the outcome of the round was provided, which included the chosen figure, the message sent by the sender (only revealed to the sender in the robustness treatments), the message received by the receiver (only revealed to the receiver in the robustness treatments), the figure indicated by the receiver and whether or not the communication is successful.[22]

To ensure subjects' understanding of the instructions, we conducted two practice rounds before the official rounds in each session. For the cash payments, we paid HKD (3 × the total number of rounds with successful communication), plus a HKD 30 show-up fee. Each session lasted for approximately one and a half hours, and subjects on average earned HKD 171.3 (USD 22.1), including the show-up fee.

4 Experimental Results

Communication was considered to be successful if the figure chosen by the receiver subject matched the actual figure chosen by the computer. We observed that languages facilitating successful communication emerged and were sustained in all treatments. The average frequency of successful communication aggregated for all official rounds, including initial rounds in which the communication success probability should have been low, was greater than 0.90 for all treatments. Figure 1 reports round-by-round time trends for the average rates of successful communication aggregated across all subjects for each treatment. The increasing trends in these figures show how communication occurred and evolved over time. The frequencies of successful communication increased, the volatility observed in the initial rounds was alleviated over time, and the trend became more stable after a few rounds.

To determine whether a common language was developed by subjects endogenously, we focused on the state-message mappings in the last 30 rounds, when behavior was more converged, as shown in Figure 1. The language mapping was precisely identified based on the collection of message-state pairs with communication success in the last 30 rounds of play. After identifying the language mapping, we checked the number of violations for each group in the last 30 rounds. It was observed that only 7 groups out of 61 had a positive number of violations, and 4 of them had only one instance of violation. We also observed that the longest message sent by the sender contained 19 symbols in the baseline treatments and very few messages had the length of 20 in the robustness treatments, implying that our length limit

[22]Note that the message sent by the sender and the message received by the receiver were the same in the baseline treatments.

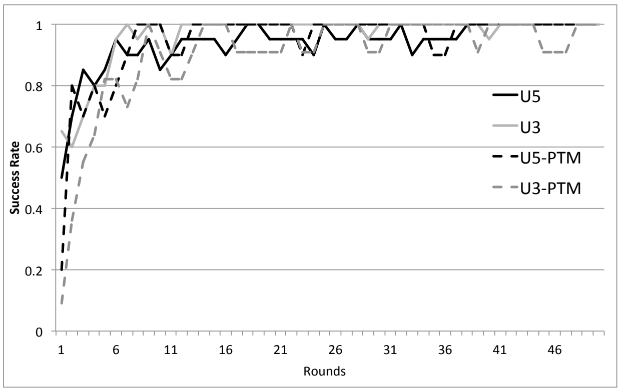


Figure 1: Time Trends – Success Rate

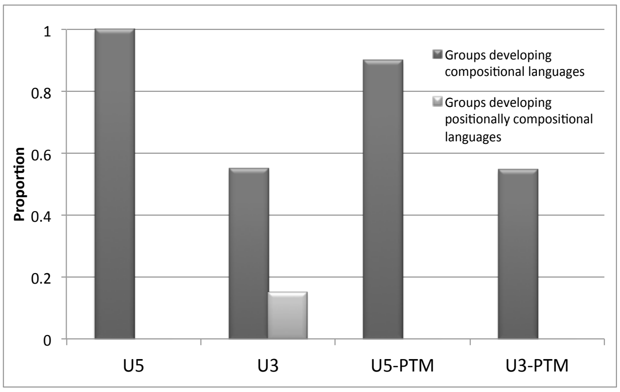


Figure 2: Compositional Languages

of up to 20 characters was almost not a binding constraint.[23]

We now present a few examples of compositional languages we identified in our experiments. The details of the languages observed in the last 30 rounds for each group are presented in Tables 4-9. In the treatment $U5$, Groups 5 and 6 in Session 1 developed the following language:

$$@ \to \bigcirc, \quad \% \to \triangle, \quad \# \to \square, \quad None \to H, \quad ! \to V,$$

and Group 7 in Session 1 developed the following language:

$$\% \to \bigcirc, \quad \$ \to \triangle, \quad \# \to \square, \quad ! \to H, \quad @ \to V,$$

where H and V represent horizontal and vertical orderings, respectively. The same kind of languages were also observed in Groups 2, 3 and 4 in Session 2 of the treatment $U5$-PTM. In the treatment $U3$, in which only three symbols were available, Group 2 in Session 2 developed the following language:

$$\$ \to \bigcirc, \quad ! \to \triangle, \quad \% \to \square, \quad !!!!!!!!!! \to H, \quad None \to V.$$

The same kind of languages were also observed in Group 4 in Session 1 and Groups 1 and 6 in Session 2 of the treatment $U5$-PTM as well as in Group 1 in Session 2 of the treatment $U3$-PTM.

The languages presented in these examples are apparently compositional and good representatives of the languages that emerged in our data. Figure 2 reports the proportions of

[23] The length of a message was quite stable across rounds after the initial 3-5 rounds. The median length of a message was 3 in the treatment $U5$ and 4 in the treatment $U3$, with the difference being statistically significant (Mann-Whitney test, p-value < 0.01). As discussed later in this section, the difference in length of messages between the two treatments was associated with the emergence of a particular type of language structures observed only in the treatment $U3$ that relies on the repetition of symbols. Such types of language structures were also observed in both treatments $U5$-PTM and $U3$-PTM. As a result, the difference in the length of a message between the two treatments becomes statistically insignificant, and the median length of messages was 4 in both treatments $U5$-PTM and $U3$-PTM (Mann-Whitney test, p-value > 0.1).

Table 4: Language in Treatment $U5$ – Session 1 – Last 30 Periods

Group	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
◯	@	@	@	@	@	@	%	@	@	@
△	!	$	!!!	!	%	%	$	!!!	$	$
□	#	#	!!!!	#	#	#	#	!!!!	#	#
Horizontal Relation	%	!	None	$	None	None	!	#	None	%
Vertical Relation	$	%	%	%	!	!	@	$	!	!
Relation Symbol Position	m	m	m	m	m	m	b	m	m	m
Unused Symbols			$		$	$		%	%	
Compositionality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of Violations	0	0	0	0	0	0	0	0	0	0
No. of Failures	0	0	0	0	0	0	0	0	0	0
Critical Moment	2	1	1	1	1	1	4	4	1	2

Note 1. Relation Symbol Position means the position of the symbol(s) indicating Relation in the message. In Relation Symbol Position, b, m, and e refer to "beginning", "middle", and "end", respectively. No. of Violations means the number of observations that violate the codes expressed above. No. of Failures means the number of observations that have unsuccessful communication. Critical Moment is defined in Definition 2. The same applies to Tables 5-9.

Note 2. In G3, # is used for separating symbols to avoid ambiguity, e.g., !!!!#!!! means □△. In G4, the relation symbol were repeated for many times in a message in the last 6 rounds, e.g. @$$$$$$$$$$$$$ represents ◯△. Since these were well understood by the receiver, we do not consider them as violations.

Table 5: Language in Treatment $U5$ – Session 2 – Last 30 Periods

Group	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
◯	!	@	@	@@	@	@	@	@	@	@
△	#	!	$	!!! or @	!!!	!	%	!	$	!
□	@	#	#	!!!! or @@@	!!	#	#	#	#	#
Horizontal Relation	%	%	%	None	None	None	None	None	None	%
Vertical Relation	$	$	!	$	%	many %	!	%	%	$
Relation Symbol Position	m	m	m	m	m	e	m	m	m	m
Unused Symbols				%	#		$	$	!	
Compositionality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of Violations	0	0	0	3	1	12	0	1	0	0
No. of Failures	0	0	1	2	1	11	0	1	0	0
Critical Moment	1	1	3	16	6	16	1	5	1	1

Note 1: In G5, symbol $ is used for separating symbols to avoid ambiguity, e.g., !!!$!! means △□. In G6, the sender kept violating the "grammar" leading to a high rate of communication failure.

Table 6: Language in Treatment $U3$ – Session 1 – Last 30 Periods

Group	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
○	%	%	$	%	$$$	$	$$	$	!	%
△	!	!	$$$	!	$$	$$$	%%	%	%	$
□	$	$	$$$$	$	$	$$$$	!!	!	$	!
Horizontal Relation	None	SR3	!	None	!	!	%%	%	SR3	None
Vertical Relation	SR3	SR2	%	SR2	!!	%	%	!	SR2	SR2
Relation Symbol Position			m		m	m	m	e		
Unused Symbols					%					
Compositionality			Yes		Yes	Yes	Yes	Yes		
Positional Compositionality							Yes	Yes		
No. of Violations	0	0	0	0	0	0	0	0	0	0
No. of Failures	0	0	0	0	0	0	0	0	0	0
Critical Moment	1	1	3	2	1	1	1	5	2	1

Note 1. SR2 and SR3 refer to “Each symbol repeated twice” and “Each symbol repeated three times”, respectively. The same applies to Tables 7-9.

Table 7: Language in Treatment $U3$ – Session 2 – Last 30 Periods

Group	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
○	%	$	$	!!!!	%	$	%	!	$	$
△	$	!	$$$	!!	!	!	$	$	$$$	%
□	!	%	$$$$	!	$	%	!	%	$$$$	!
Horizontal Relation	None	!!!!!!!!!!	%	%	%	$$$ or %%% or !!!	None	SR3	!!	See Note 1
Vertical Relation	SR2	None	!	$	!	$$$$$$ or %%%%%% or !!!!!!	SR2	SR2	!	See Note 1
Relation Symbol Position		m	m	m	m	m			m	
Unused Symbols									%	
Compositionality		Yes	Yes	Yes	Yes	Yes			Yes	
Positional Compositionality					Yes					
No. of Violations	0	0	0	0	0	0	0	0	0	0
No. of Failures	0	0	0	0	0	0	0	0	0	1
Critical Moment	3	1	1	4	1	5	1	5	6	9

Note 1. In $G10$, the horizontal relation is represented by the three-times repetition of the symbol representing the object in Ω_3. The length of messages for the state with horizontal relation is thus 4. The vertical relation is represented by four-times repetition of the symbol representing the object in Ω_1 and two-times repetition of the symbol representing the object in Ω_3. The length of messages for the state with vertical relation is 6.

Table 8: Language in Treatment *U*5-*PTM* – Sessions 1-2 – Last 30 Periods

Group	S1G1	S1G2	S1G3	S1G4	S2G1	S2G2	S2G3	S2G4	S2G5	S2G6
◯	@@@	#####	@	@	@	@	@	%	#	$
△	!!!	!!!	!!!	%	%	%	$	#	!	#
□	###	@@@@	####	#	#	#	#	$	@	@
Horizontal Relation	%	%	None	$$$$$$$$$$$$$$$$$$	None	$	None	!	None	%%
Vertical Relation	None	$	$...%	!!!	!!!!!	!	%	@	SR2	!!
Relation Symbol Position	m	e	See Note 1	e	m	e	e	b		e
Unused Symbols	$				$		!		%,$	
Compositionality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
No. of Violations	1	0	0	0	0	0	0	0	0	0 (See Note 2)
No. of Failures	2	0	0	0	0	0	0	0	1	0
Critical Moment	14	1	7	2	2	2	3	5	2	1

Note 1. In S1G3, $ in the beginning with % in the middle separating symbols indicating the two shapes is used to indicate vertical relation.
Note 2. In S2G6, in the last period, !! was repeated many times but it did not jeopardize the success of communication. Therefore, we do not consider it as a violation.

Table 9: Language in Treatment *U*3-*PTM* – Sessions 1-2 – Last 30 Periods

Group	S1G1	S1G2	S1G3	S1G4	S2G1	S2G2	S2G3	S2G4	S2G5	S2G6	S2G7
◯	!!	%	%	%	$	!	!	$	$	$	!$
△	$$	!	$	$	!	$	!!!	%	!	!!!	!%
□	!$	$	!	!	%	%	!!!!	$%	%	%%%%	!!
Horizontal Relation	None	None	SR2	None	%%%	None	%	!!	None	None	None
Vertical Relation	%	See Note 1	SR3	SR2	$$$	SR2	$	!	SR2	$ or % or !	!
Relation Symbol Position	e				e		m	b		m	m
Unused Symbols											
Compositionality	Yes				Yes		Yes	Yes		Yes	Yes
No. of Violations	0	0	2	0	0	0	0	0	0	0	1
No. of Failures	0	4	2	0	0	0	0	1	0	1	1
Critical Moment	13	18	13	7	2	2	4	6	1	9	3

Note 1. In S1G2, the symbol representing the second shape was repeated for many (≥ 10) times to represent the vertical relation, where the numbers of repetitions vary for different figures.

groups with compositional grammars in the last 30 rounds for each treatment. In the treatment $U5$, 20 out of the 20 groups developed compositional grammars, whereas 11 out of the 20 groups in the treatment $U3$ developed compositional grammars. Similarly, in the treatment $U5$-PTM, 9 out of the 10 groups developed compositional grammars, whereas 6 out of the 11 groups in the treatment $U3$-PTM developed compositional grammars. Thus, we have the following result confirming our Hypothesis 1.

Result 1 (Compositionality) *In all of the four treatments, compositional languages frequently emerged. In the treatments $U5$ and $U5$-PTM, more than 90% of groups developed compositional languages. In the treatments $U3$ and $U3$-PTM, more than 50% of groups developed compositional languages.*

Figure 2 also indicates that 3 out of the 20 groups in the treatment $U3$ (Groups 7 and 8 in Session 1, Group 5 in Session 2) developed positionally compositional grammar, but no such grammar was observed in the treatment $U5$.[24] For example, Group 8 in Session 1 of the treatment $U3$ developed the following language:

$$\$\_\_ \rightarrow \bigcirc, \%\_\_ \rightarrow \triangle, !\_\_ \rightarrow \square, \_\$\_ \rightarrow \bigcirc, \_\%\_ \rightarrow \triangle, \_!\_ \rightarrow \square, \_\_\% \rightarrow H, \_\_! \rightarrow V,$$

where, for example, $!\_\_$, $\_!\_$ and $\_\_!$ represent the symbol ! positioned in the first, second, and third place, respectively, of a message string with a length of three.

Fisher's exact test reveals that compositional grammar was significantly more likely to emerge in the treatment $U5$ ($U5$-PTM) than in the treatment $U3$ ($U3$-PTM) (p-values < 0.01); among groups with compositional grammar, positionally compositional grammar was significantly more likely to emerge in the treatment $U3$ than in the treatment $U5$ (p-value < 0.05). However, no such positionally compositional grammars were observed in any of the robustness treatments. We summarize this finding as follows.

Result 2 (Positional Compositionality) *Compositional grammars emerged more frequently in the treatment $U5$ ($U5$-PTM) than in the treatment $U3$ ($U3$-PTM). Among groups with compositional grammars, positionally compositional grammars emerged more frequently in the treatment $U3$ than in the treatment $U5$. However, no such difference existed between the treatments $U5$-PTM and $U3$-PTM.*

Result 2 lends support to our Hypothesis 2 for the baseline treatments. Note that a compositional grammar in the treatment $U5$ ($U5$-PTM) is straightforward because one distinct symbol could be used to represent one factor (shape or relation) of the state space. In contrast, using one symbol to represent one factor in the state space requires the use of positional compositionality in the treatment $U3$ ($U3$-PTM). Otherwise, to achieve efficient communication,

[24]Other than the three groups developing positionally compositional grammars in the treatment $U3$, the compositional grammars developed in all other groups satisfy the *simple* compositionality discussed in Footnote 16.

individuals must be more creative in composing a message. For example, we observed that Group 6 of Session 1 in the treatment $U3$ developed the following language:

$$\$ \to \bigcirc, \quad \$\$\$ \to \triangle, \quad \$\$\$\$ \to \square, \quad ! \to H, \quad \% \to V.$$

Similar languages were observed in Groups 3 and 5 of Session 1 and Groups 3, 4, and 9 of Session 2 in the treatment $U3$ as well as in Groups 3, and 4 of Session 2 in the treatment $U3$-PTM. These languages are still compositional albeit not positionally compositional.

Our experiments clearly indicated that learning was based on the structure of languages. The average frequency of successful communication aggregated for all 50 rounds was surprisingly high, 0.939 and 0.948 for the treatments $U5$ and $U5$-PTM, and 0.965 and 0.902 for the treatments $U3$ and $U3$-PTM, respectively. To measure how many *distinct* observations were required for each group to learn the language, we define the *Critical Moment* (CM) for each group as follows:

Definition 2 (Critical Moment)

$$CM = ER - RS$$

where ER is the earliest round in which communication is successful and after which communication is successful in all but at most one round, and RS is the number of repetitive states realized before ER.

Three remarks about the critical moment are in order. First, the definition of RS is taken to rule out the observations with repetitive states, which are either indicated by the same messages as before, in which case nothing new is learned, or indicated by different messages, in which case learning is confounded.[25] Second, $CM = k$ implies that $(k - 1)$ distinct observations were required to learn the language. Third, our definition of CM is tolerant to one communication error after the period of ER.[26] The critical moment for each group is reported in Tables 4-9.

Hypothesis 3 implies that CM for the groups developing compositional grammars should be no larger than 4. Moreover, the possible focal associations in the baseline treatments suggests that CM for such groups could be even strictly lower than 4. We find that the average CMs of the groups developing compositional grammars in the treatments $U5$ and $U3$ are respectively 3.5 and 2.6, which are significantly smaller than 4 (Wilcoxon signed-rank tests, p-value $= 0.036$ and 0.051, respectively). The average CMs of the groups developing compositional grammars in the treatments $U5$-PTM and $U3$-PTM are respectively 4.1 and

[25] For example, Group 5 in Session 2 of the treatment $U5$ had $ER = 9$, but there were three repetitive states realized (in Rounds 2, 6, and 7) before Round 9 so that $RS = 3$. As a result, CM of the group became 6.

[26] For example, Group 3 in Session 2 of the treatment $U5$ and Group 7 in Session 2 of the treatment $U3$-PTM have had successful communication after having two distinct observations, except for Round 33 and Round 47, respectively. According to our definition, the CM for both groups is 3 (not 34 and 48).

6.2, which are not statistically different from 4 (Wilcoxon signed-rank tests, p-values > 0.1). These observations allow us to confirm our Hypothesis 3. We also find that CM is not higher than four in 16 out of the 20 groups in the treatment $U5$, in 7 out of the 10 groups in the treatment $U5$-PTM, in 15 out of the 20 groups in the treatment $U3$, and in 5 out of the 11 groups in the treatment $U3$-PTM. This result implies that only three or fewer distinct observations were needed for these groups to learn their languages, an impossible scenario unless they relied on the structure of language to learn because the number of states in our experiments was 18.

Result 3 (Learning Efficiency of Compositional Languages) *In all of the four treatments, the groups developing compositional grammars were able to learn the languages based on weakly fewer than three distinct observations on average. Moreover, a majority of groups learned their languages from weakly fewer than three distinct observations, which implies that learning is based on the structure of languages.*

Although the theory predicts that compositional grammar optimizes learning efficiency, all groups without compositional grammars in our experiments developed language structures that also facilitated communication. For instance, Group 1 of Session 1 in the treatment $U3$ developed the following mapping:

$$\%! \to \bigcirc\triangle, \%\%\%!!! \to \begin{array}{c}\bigcirc\\ \triangle\end{array}, \%\$ \to \bigcirc\square, \%\%\%\$\$\$ \to \begin{array}{c}\bigcirc\\ \square\end{array}, !\% \to \triangle\bigcirc, !!!\%\%\% \to \begin{array}{c}\triangle\\ \bigcirc\end{array}, \ldots \quad (2)$$

In this mapping, each component of Ω_1 and Ω_3 is represented by a distinct symbol ($\% \to \bigcirc$, $! \to \triangle$, $\$ \to \square$) while repeating each symbol three times indicates the vertical relation. Adding a new meaning by repeating the same symbol multiple times is somewhat similar to an *inflection* in natural languages, a syntactical rule that carries meaning. A similar language was observed in Groups 2, 4, 9 and 10 of Session 1 and Groups 1, 7, 8 and 10 of Session 2 in the treatment $U3$ as well as in Group 5 of Session 2 in the treatment $U5$-PTM, Groups 2, 3, 4 of Session 1 and Groups 2 and 5 of Session 2 in the treatment $U3$-PTM. According to Definition 1, these languages are *not* compositional.[27] However, they are efficient in learning.[28]

In the treatment $U3$, for which 11 groups developed compositional grammars, although the average CM conditional on the development of compositional languages was 2.6, which is less than 2.8, the average CM conditional on the development of non-compositional languages,

[27] These languages might be regarded as compositional according to the alternative definitions introduced in Footnote 12 because "the meanings of its composite expressions are systematically derived from the meanings of their parts and the way in which these parts are combined." If we think of these languages as compositional, then the percentage of groups developing compositional languages is 100% for the treatments $U3$, $U5$-PTM, and $U3$-PTM. However, whether regular inflections are compositional or not is an old debate. See Pinker (1999) for more details.

[28] Taking the mapping presented in (2) for example, at most three distinct observations (e.g. ($\bigcirc\bigcirc$, %%), ($\begin{array}{c}\bigcirc\\ \bigcirc\end{array}$, %%%%%%), ($\bigcirc\triangle$, %!)) are sufficient for learning.

the difference was not statistically significant (Mann-Whitney test, p-value > 0.1). Similarly, in the treatment $U3$-PTM, the difference between the CM conditional on the development of compositional languages and that conditional on the development of non-compositional languages was neither substantial (on average 6.2 vs. 8.2) nor statistically significant (Mann-Whitney test, p-value > 0.1). Consequently, in the treatments $U3$ and $U3$-PTM, there was no significant difference in the success rate of communication between the groups developing compositional languages and groups without such languages either in the early stages of the experiment (in the first five periods) or throughout the entire experiment (Mann-Whitney tests, p-values > 0.1).[29]

With the abovementioned non-compositional but (learning-)efficient languages, the treatment $U3$ did not underperform the treatment $U5$ in communication. The Mann-Whitney test shows that there was no statistically significant difference in terms of success rate or CM between the treatments $U5$ and $U3$ (p-values > 0.1). Similarly, there was no significant difference between the treatments $U3$-PTM and $U5$-PTM in terms of the success rate or CM (p-values > 0.1). Therefore, in spite of the scarcity of symbols in the repertoire in the treatment $U3$ ($U3$-PTM), which likely caused the lower adherence of compositional grammars, the subjects in the treatment $U3$ ($U3$-PTM) were still able to develop languages as effective as those in the treatment $U5$ ($U5$-PTM) to facilitate learning and communication. This is in contrast to Selten and Warglien's (2007) result that less permissible symbols were an obstacle to the emergence of common language codes, where the use of symbols was costly. The costless symbols in our setup allowed players to repeat permissible symbols in an innovative way to communicate efficiently. Rejecting Hypothesis 4, we summarize these results as follows.

Result 4 (No Efficiency Gain from More Permissible Symbols) *There was no significant difference between the treatments* $U5$ *(*$U3$*-PTM) and* $U3$ *(*$U5$*-PTM) either in terms of the frequency of successful communication or in terms of CM.*

Tables 4-9 show that the mappings from the shapes to the symbols varied across groups. As expected, however, we also found that some mappings were more common than others in the baseline treatments. For example, @ and % were often used to represent ◯, and # was frequently used to represent □ in the treatment $U5$, while % was often used to represent ◯ in the treatment $U3$. This pattern did not disappear under the privatized translation of messages protocol adopted in the treatments $U5$-PTM and $U3$-PTM.[30] While the receivers in the baseline treatments directly observed these associations from the feedback of the outcome in each round, such associations were destroyed on the receiver's side by the PTM protocol in the robustness treatments.

[29]The same analysis could not be done for the treatments $U5$ and $U5$-PTM because we observed that all 20 groups in $U5$ and 9 groups out of 10 in $U5$-PTM developed compositional languages.

[30]It thus reveals that senders chose particular symbols not just because they believe the choice will help receivers understand the language mapping, but because it is easier / more natural for themselves to use the symbols.

If the focal associations speed up learning while the *PTM* protocol eliminates the effect of these focal associations, one should expect to observe that (1) in the treatment $U5$ ($U3$) the groups with languages relying on the focal associations have lower *CM* than those without, and (2) the groups with languages relying on the focal associations have lower *CM* in the treatment $U5$ ($U3$) than in the treatment $U5$-*PTM* ($U3$-*PTM*). We found that focal associations indeed accelerate learning when the associations were not destroyed by the *PTM* protocol. The effect seems stronger when there were less permissible symbols, in which case the senders were required to come up with some clever but relatively more complicated mapping.

More precisely, the groups with the focal association (%→ ◯) had significantly lower *CM* than the groups without (1.4 vs. 3.4; Mann-Whitney test, p-value = 0.07) in the treatment $U3$.[31] Meanwhile, in the treatments with 5 symbols, the groups with any of the focal associations (@, %→ ◯, #→ □) had lower *CM* in the treatment $U5$ than in the treatment $U5$-*PTM*, although the difference was only marginally significant (3.6 vs. 5; Mann-Whitney test, p-value = 0.09); in the treatments with 3 symbols, however, the groups with the focal association (%→ ◯) had lower *CM* in the treatment $U3$ than in the treatment $U3$-*PTM*, and the difference is statistically significant and substantial (1.4 vs. 12.7; Mann-Whitney test, p-value = 0.01). In contrast, there was no statistical difference in *CM* for groups without the focal associations between the treatments $U5$ and $U5$-*PTM*, or between the treatments $U3$ and $U3$-*PTM*.

These results were translated into the following additional findings for the treatments $U5$ ($U3$) and $U5$-*PTM* ($U3$-*PTM*). There was no significant difference between the treatments $U5$ and $U5$-*PTM* either in terms of *CM* or in terms of the frequency of successful communications (Mann-Whitney tests, p-values > 0.1). On the contrary, the treatment $U3$-*PTM* presented a few noticeable differences from the treatment $U3$. First, *CM* from the treatment $U3$-*PTM* was significantly higher than that from the treatment $U3$ (7.1 vs. 2.7; Mann-Whitney test, p-value = 0.01). Second, the average success rates over the first 5 rounds were 0.71 and 0.6 respectively for the treatments $U3$ and $U3$-*PTM*, with the difference being statistically significant (Mann-Whitney test, p-value = 0.079). This difference did not disappear even when we took the frequency of successful communication from all 50 rounds (Mann-Whitney test, p-value = 0.019). The difference in magnitude was, however, not very substantial as the success rates over all 50 rounds were 0.948 and 0.902 in the treatments $U3$ and $U3$-*PTM*, respectively. Finally, Fisher's exact tests reveal that there is no significant difference in terms of the frequency of compositional languages between the treatments $U5$ and $U5$-*PTM* or between the treatments $U3$ and $U3$-*PTM* (p-values > 0.3).

Result 5 (Effect of Privatized Translation of Messages) *There was no significant difference between the treatments* $U5$ *and* $U5$-*PTM either in terms of the frequency of successful communication or in terms of CM. However, CM was significantly higher and the frequency*

[31]We could not conduct the same test for treatment $U5$ because there was only one group that did not use any of the focal associations mentioned above in the treatment.

of successful communication was significantly lower in the treatment $U3$-PTM than in the treatment $U3$.

It is worth discussing a few other interesting features observed in the artificial codes identified in our experiments. First, in the baseline treatments, most groups with compositional languages positioned the symbol representing the relation in the middle of a message string. Only two groups put the symbol representing the spatial relation at the end of a message string, whereas there was one group that put the symbol representing the spatial relation at the beginning of a message string. In the robustness treatments, however, seven groups with compositional languages positioned the symbol representing the spatial relation at the end of a message string, whereas there were five and two groups that put the symbol representing the spatial relation respectively in the middle and at the beginning of a message string.

Second, compositional grammars saved the use of additional symbols even if using them was costless. In ten groups in the treatment $U5$, two groups in the treatment $U3$, and four groups in the treatment $U5$-*PTM*, one or two symbols were never used. All of these groups developed compositional grammar except for one.

5 Concluding Remarks

Our study experimentally examined the structures of language in artificial codes. We focused on one key property of language: compositionality. The results of our laboratory experiment reveal that compositional grammars in artificial codes emerged and were sustained. Subjects in the laboratory developed grammatical common codes that facilitated efficiency in learning languages. Notably, our work is among the first empirical investigations of common linguistic properties and the emergence of simple grammar-like structures of artificial languages in the laboratory.

Three possible extensions of our study may be worth mentioning. First, the assumption of a conflict of interest among economic players plays a key role in modern economic theory. One may go one step further to investigate how a conflict of interest between the sender and receiver *á la* Crawford and Sobel (1982) can generate additional insights into experimental semiotics. Second, it would also be interesting to determine whether our results regarding the use of artificial codes would extend to a less abstract design as in the work of Weber and Camerer (2003), who studied corporate culture embedded in the particular language shared between players in an organization. Third, while we follow the main body of the experimental literature to consider a one-sender-one-receiver environment, the emergence of language in a larger group context would be also worth investigating.

References

[1] Blume, A. 2000. "Coordination and Learning with a Partial Language." *Journal of Economic Theory*, 95: 1-36.

[2] Blume, A. 2005. "A Learning-Efficiency Explanation of Structure in Language." *Theory and Decision*, 57: 265-285.

[3] Blume, A., D. DeJong, Y. Kim and G. Sprinkle. 1998. "Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games." *American Economic Review*, 88 (5): 1323-1340.

[4] Bruner, J., C. O'Connor, H. Rubin, and S. M. Huttegger. 2014, "David Lewis in the Lab: Experimental Results on the Emergence of Meaning." *Synthese*, doi: 10.1007/s11229-014-0535-x.

[5] Crawford, V. and H. Haller. 1990. "Learning How to Cooperate: Optimal Play in Repeated Coordination Games." *Econometrica,* 58: 571-595.

[6] Crawford, V. and J. Sobel. 1982. "Strategic Information Transmission." *Econometrica,* 50: 1431-1451.

[7] Fischbacher, U. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics*, 10(2), 171-178.

[8] Franke, M., 2014. "Creative Compositionality from Reinforcement Learning in Signaling Games." *The Evolution of Language*: Proceedings of the 10th International Conference (Evolang 10). Ed. by Erica A. Cartmill et al. Singapore: World Scientific, pp. 82-89.

[9] Franke, M., 2016. " The Evolution of Compositionality in Signaling Games." *Journal of Logic, Language and Information*, 25: 355-377.

[10] Frege, G. 1882. "Zeitschrift fuer Philosophie und Philosophische." *Kritik*, 100: 25-30.

[11] Galantucci, B. and S. Garrod. 2010. "Experimental Semiotics: A New Approach for Studying the Emergence and the Evolution of Human Communication." *Interaction Studies*, 11: 1-13.

[12] Galantucci, B. and S. Garrod. 2011. "Experimental Semiotics: a Review." *Frontiers in Human Neuroscience*, 17: 1-15.

[13] Hong, F., W. Lim, and X. Zhao. 2015. "The Emergence of Language Structures in Artificial Codes." mimeo.

[14] Janssen, T. M. V., "Montague Semantics", *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2016/entries/montague-semantics/ >.

[15] Kirby, S., H. Cornish, and K. Smith. 2008. "Cumulative Cultural Evolution in the Laboratory: An Experimental Approach to the Origins of Structure in Human Language." *PNAS*, 105: 10681-10686.

[16] Lewis, D., 1969. *Convention,* Cambridge, MA: Harvard University Press.

[17] Lipman, B. 2003. "Language and Economics." N. Dimitri, M. Basili and I. Gilboa, eds., *Cognitive Processes and Rationality in Economics*, London: Routledge.

[18] Pinker, S. 1999. "Words and Rules: The Ingredients of Language." New York, NY: Harper Perennial.

[19] Rubinstein, A. 1996. "Why are Certain Properties of Binary Relations Relatively More Common in Natural Language?" *Econometrica*, 64: 343-356.

[20] Rubinstein, A. 2000. "Economics and Language." Cambridge: Cambridge University Press.

[21] Szabó, Z. G., "Compositionality", *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2013/entries/compositionality/ >.

[22] Selten, R., and M. Warglien. 2007. "The Emergence of Simple Languages in an Experimental Coordination Game." *Proceedings of the National Academy of Sciences of the United States of America*, 104(18): 7361-7366.

[23] Steinert-Threlkeld, S., 2016. "Compositional Signaling in a Complex World." *Journal of Logic, Language and Information*, forthcoming.

[24] Weber, R.A., and C. Camerer. 2003. "Cultural Conflict and Merger Failure: An Experimental Approach." *Management Science*, 49: 400–415.

Appendix A. Experimental Instructions – Treatment $U5$-$\boldsymbol{PTM}$

INSTRUCTION

Welcome to the experiment. This experiment studies decision making between individuals. In the following two hours, you will participate in 50 rounds of decision making. Please read the instructions below carefully; the cash payment you will receive at the end of the experiment may depend on how well you make your decisions according to these instructions. Please turn off your mobile phone and any other electronic devices. Communication of any kind with other participants is not allowed.

Your Role and Decision Group

Half of the participants will be randomly assigned the role of Member A and the other half the role of Member B. Your role will remain fixed throughout the experiment. Prior to the first round, one Member A will be paired with one Member B to form a group of two. The group formation will remain fixed for all 50 rounds. That is, you will interact with only one participant throughout the entire study. The two members in a group make decisions that will affect their rewards in the round.

Your Decision in Each Round

You will be presented with a box having eighteen figures on your screen as you see in Figure 3(a) or 4(a). The following table shows these eighteen figures.

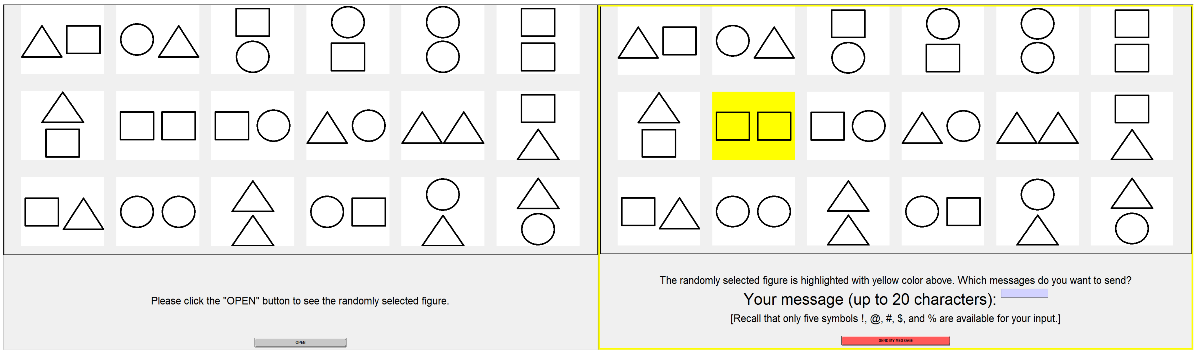
○ ○	△□	□ □	△ □	△ △	○△	□□	○ □	□ △
△○	○□	△△	○ △	△ ○	○○	□△	□ ○	□○

Note: The order of figures in this table is randomly determined for each participant.

Table 10: Composition of Figures

In each round and for each individual, the positions of these figures on your computer monitor are randomly determined. In each round and for each group, the computer will randomly select one figure in the box. Each figure in the box has an equal chance to be selected. The selected figure will be revealed to Member A only. Member B, without seeing the selected figure, will have to guess what the figure is.

Member A's Decision



(a) Member A's Screen: Before "Open" (b) Member A's Screen: After "Open"

Figure 3: Screen Shots

In each round, you will be presented with a box having the eighteen figures on your screen as explained above. The positions of the eighteen figures are randomly determined for each individual in each round. If you click the "open" button at the bottom of your screen, then you will see that one of the figures is highlighted with yellow color on your screen (see Figure 3(b)). The highlighted figure is the randomly selected figure for your group.

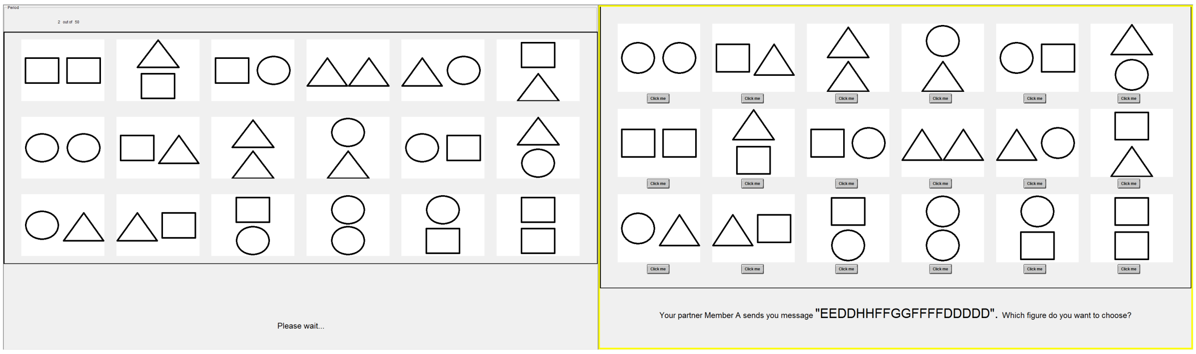
After seeing the randomly selected figure, you will be asked to send a message consisting of a string (with an arbitrary length up to 20 characters) of available symbols !, @, #, $ and %, as presented in the Figure 3(b). You need to input your message using the symbols available in your keyboard. You cannot send any message that involves any symbol(s) other than the five specified ones above. Once you finish inputting the message, you need to click the "SEND MY MESSAGE" button. Then your decision in the round is completed.

Privatized Message Translation and Transmission

After Member A inputs his/her message and clicks the "SEND MY MESSAGE" button, each distinct symbol in the written message is translated into one distinct letter among D, E, F, G and H. Which symbol is translated into which letter is *randomly* determined at the beginning of the experiment and will not be revealed to Member A and Member B. But, once uniquely determined, **the same symbol-letter translation will be used throughout the 50 rounds of interactions for each group**. The translated message will be transmitted to the paired Member B, who will then be asked to guess the randomly selected figure.

Member B's Decision

In each round, you will be presented with a box having eighteen figures on your screen as explained above. You will be waiting for the message from your paired Member A. Once your paired member A finishes making his/her message choice, you will see a translated version of a message from Member A on your screen. Then you will be asked to choose one figure out of the eighteen figures presented in Table 10.



(a) Member B's Screen: Waiting (b) Member B's Screen: Figure Choices

Figure 4: Screen Shots

Here are more details about the procedure for your figure choice. In your screen (see figure 4(b)), you will be presented with the eighteen figures, each of which has the "Click me" button below. The positions of the eighteen figures are randomly determined for each individual in each round. You can press the "Click me" button right below each figure to choose the figure. Then your decision in the round is completed.

Your Reward in Each Round

Your reward in the experiment will be expressed in terms of experimental currency unit (ECU). In each round, you will receive 3 ECU if the randomly selected figure is matched with the figure chosen by Member B in your group. If there is a mismatch between the two figures, then you will receive 0 ECU.

Information Feedback

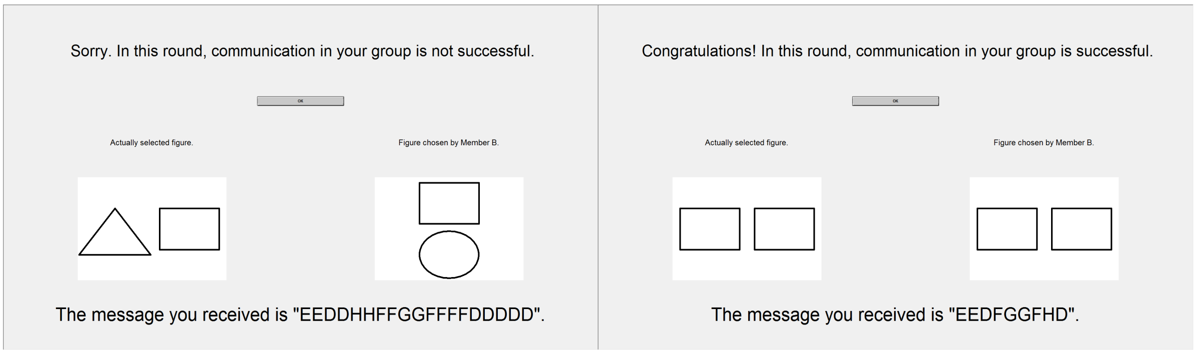
At the end of each round, the computer will provide a summary for the round: which figure was selected and revealed to Member A, the message written by Member A (only revealed to Member A) and its translated version (only revealed to Member B), Member B's figure choice, whether or not your group has successful communication in the round, and your total earning in the ECU term. Figure 5 below is an example of the screen for the information feedback.

Your Cash Payment

At the end of the experiment, the sum of ECU you earned in all 50 official rounds will be converted into cash at an exchange rate of HK$1 per ECU. The cash payment at the end of the study will be this cash amount plus a HK$30 show-up fee. Precisely,

Your total cash payment = HK$ [The sum of ECU in all 50 rounds + 30]

Practice Rounds



(a) Outcome Screen: Failure (b) Outcome Screen: Success

Figure 5: Screen Shots

To ensure your understanding of the instructions, we will provide you with 2 practice rounds. The practice rounds are part of the instructions which are not relevant to your cash payment; its objective is to get you familiar with the computer interface and the flow of the decisions in each round. Once the practice rounds are over, the computer will tell you "The official rounds begin now!"

Administration

Your decisions as well as your monetary payment will be kept confidential. Remember that you have to make your decisions entirely on your own; please do not discuss your decisions with any other participants. Upon finishing the experiment, you will receive your cash payment. You will be asked to sign your name to acknowledge your receipt of the payment. You are then free to leave. If you have any question, please raise your hand now. We will answer your question individually. Otherwise, we will proceed to the practice rounds.