 doi:10.20944/preprints202210.0399.v1

Article

EXP-SC: A Semantic Communication Model Based on Information Framework Expansion and Knowledge Collision

Gangtao Xin [1,2] and Pingyi Fan [1,2,*]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China; xgt19@mails.tsinghua.edu.cn
[2] Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China;
* Correspondence: fpy@tsinghua.edu.cn; Tel.: +86-010-6279-6973

Abstract: Semantic communication is not obsessed with improving the accuracy of transmitted symbols, but is concerned with expressing the desired meaning that the symbol sequence exactly carried. However, the generation and measurement of semantic messages are still an open problem. Expansion combines simple things into complex systems and even generates intelligence, which is consistent with the evolution of the human language system. We apply this idea to semantic communication system, quantifying and transmitting semantics by symbol sequences, and investigate the semantic information system in a similar way as Shannon did for digital communication systems. This work was the first to propose the concept of semantic expansion and knowledge collision, which may provide a new paradigm for semantic communications. We believe that expansion and collision will be the cornerstone of semantic information theory.

Keywords: semantic information theory; semantic communications; information theory; 6G; game theory

1. Introduction and Overview

Shannon's information theory [1] answers two fundamental questions in digital communication theory [2]: the ultimate data compression rate (answer: the entropy H) and the ultimate reliable transmission rate of communication (answer: the channel capacity C). It addresses technical problems in communication systems, enabling it to receive the same symbols as the sender. However, with the ever-increasing demand for intelligent wireless communications, the communication architecture is evolving from only focusing on technical level to intelligent interconnection of everything [3]. Weaver [4] categorized communications into three levels about 70 years ago:

Level A. How accurately can the symbols of communication be transmitted? (The technical problem.)

Level B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)

Level C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)

Recently, semantic information theory including semantic communications has attracted much attention. One reasonable and feasible measure of semantic messages is still an open problem, which may be the greatest challenge for the new developments of semantic communication (SC) systems. On the other hand, currently in 6G networks, intelligent interconnections of everything will bring the conventional paradigm to the communication mode with higher requirements of semantic interaction. In this regard, finding a way to quickly reflect the semantic processing of messages from the sender and the receiver will become one more promising pathway to the 6G intelligent networking systems.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0399.v1

This paper proposes a new communication system based on information framework expansion and knowledge collision, by extending Shannon's theory of communication (Level A) to a theory of semantic communication (Level B). Our work is initially influenced by Jinho and Jihong [5], with new contributions in the following:

- We generalize the work of Jinho and Jihong from single fixed semantic type to semantic dynamically scaling up mode, referred to as expansion, and get the idea of knowledge collision, which can reflect the asynchronous knowledge update process of the sender and the receiver in some degree. In addition, we also take into account the effect of channel noise on the model.
- We present a new measure related to the semantic communication system based on semantic expansion and knowledge collision, called Measure of Comprehension and Interpretation, which can measure the semantic entropy of discrete sources.
- We discuss the additional gains from expansion and demonstrate that knowledge matching of its asynchronous scaling up plays a key role in semantic communications.

As a primary work, the system proposed in this paper only focuses on the discrete cases and makes some drastic simplifications. Moreover, we are not concerned with the effective problem (Lever C), which is beyond the scope of this paper. However, we believe that these simplifications are necessary for us to focus on the "core" issues, and that even this model may form a foundation for a general semantic information theory. To the best of our knowledge, this is the first study to explore the semantic expansion and asynchronous scaling up of semantic knowledge. The insights gained from this work may be of assistance to the future intelligent semantic communication system and 6G.

The rest of this paper is organized as follows. Section 2 introduces the related work of semantic communications. It summarizes key concepts and conclusions of Shannon's information theory and semantic information theory. In Section 3, we present the generalized model of semantic communication from two aspects, one is trying to setup a close relationship with Shannon communication model, and the other is trying to find the feasible modification of the model so that it can exactly reflect the quickly surging of semantic communications from the user requirements. In Section 4, we lay out the implementation details of the simulation and discuss the experimental results. Finally, we conclude the paper in Section 5.

2. Related Work

In this section, we introduce the framework of semantic information theory, especially the key concept, semantic entropy. It will then go on to the work that we are concerned with, semantic communication as a signaling game with correlated knowledge bases.

2.1. Preliminaries

Although Shannon's information did not address the semantic problem of communications, it provided important insights on the message processing techniques associated with the focus of both the sender and the receiver's attention. Thus, in this subsection, we briefly introduce the main concepts and theorems in Shannon's information theory.

Entropy. It is a measure of the uncertainty of a random variable [2]. The *entropy* $H(X)$ of a discrete random variable X with probability distribution $p(x)$, is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{1}$$

Mutual information. It is a measure of the amount of information that one random variable contains about another. Mutual information can be seen as the reduction in the uncertainty of one random variable due to the knowledge of the other. Consider two random variables X and Y with a joint probability mass function $p(x, y)$ and marginal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0399.v1

probability mass functions $p(x)$ and $p(y)$, the mutual information between X and Y is defined as

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \frac{p(x,y)}{p(x)p(y)}. \quad (2)$$

Channel capacity. The channel capacity is the maximum of mutual information with given the conditional transit probability from X to Y, $p(y|x)$. It is defined by

$$C = \max_{p(x)} I(X;Y), \quad (3)$$

where X and Y are the input and output of the channel, respectively.

Source coding theorem. As $N \to \infty$, N i.i.d. (independent identically distribution) random variables with entropy $H(X)$ can be compressed into more than $NH(X)$ bits to represent them completely, that is, the information loss can be negligible in the viewpoint of Shannon's information theory. Conversely, if they are compressed into fewer than $NH(X)$ bits to represent them, there will be a loss of information certainly.

Channel coding theorem. For a discrete memoryless channel, all rates below capacity C are achievable. Conversely, any sequence of codes with negligible error must have obey the rule, its transmission rate is not greater than the channel capacity, $R \leq C$.

2.2. Semantic Information Theory

Carnap and Bar-Hillel [6] were the first to propose the concept of semantic information theory, using logical probability rather than statistical probability to measure the semantic entropy of a sentence. It's necessary here to clarify exactly that we use p and m to denote statistical probability and logical probability in this paper, respectively. The logical probability of a sentence is measured by the likelihood that the sentence is true in all possible situations [7]. Then, the semantic information of the message e is defined as

$$H_S(e) = -\log_2(m(e)), \quad (4)$$

where $m(e)$ is the logical probability of e. However, this metric led to a paradox that any fact has an infinite amount of information when it contradicts itself, i.e., $H_S(e \wedge \neg e) = \infty$. Floridi [8] solved the paradox in Carnap and Bar-Hillel's proposal [6], adopting the relative distance of semantics to measure the amount of information.

Bao *et al.* [7] defined semantic entropy of a message x as

$$H_S(x) = -\log_2(m(x)), \quad (5)$$

where the logical probability of x is given by

$$m(x) = \frac{\mu(W_x)}{\mu(W)} = \frac{\sum_{w \in W, w \models x} \mu(w)}{\sum_{w \in W} \mu(w)} \quad (6)$$

Here, W is the symbol set of a source, $\models$ is the proposition satisfaction relation, and W_x is the set of models for x. In addition, μ is a probability measure, $\sum_{w \in W} \mu(w) = 1$.

Besides logical probability, there are some definitions of semantic entropy based on different backgrounds [9][10]. D'Alfonso [11] utilized the notion of truthlikeness to quantify semantic information. Kolchinsky and Wolpert [12] defined semantic entropy as the syntactic information that a physical system has about its environment which is causally necessary for the system to maintain its own existence. Kountouris and Pappas [13] advocated for assessing and extracting the semantic value of data at three different granularity levels: microscopic scale, mesoscopic scale and macroscopic scale.

Analogous to Shannon's information theory, some related theories at the semantic level, such as semantic channel capacity, semantic rate distortion and information bottleneck

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0399.v1

[14], are also being explored. Based on (5), Bao *et al.* further proposed *semantic channel coding theorem* [7]. For every discrete memoryless channel, the channel capacity

$$C_S = \sup_{P(X|W)} \{I(X;Y) - H_{K_S,I_S}(W|X) + \overline{H_{S;K_r,I_r}(Y)}\}, \tag{7}$$

has the following property: For any $\epsilon > 0$ and $R \leq C_S$, there is a block coding strategy such that the maximal probability of semantic error is $\leq \epsilon$. $I(X;Y)$ is the mutual information between the input X of the channel and the output Y of the channel, $H_{K_S,I_S}(W|X)$ is the equivocation of the semantic encoder, given the sender's local knowledge K_S and inference procedure I_S. $\overline{H_{S;K_r,I_r}(Y)} = \sum_y p(y)H_s(y)$ is the average information of received messages, given the receiver's local knowledge K_S and inference procedure I_S.

Moreover, H. Vincent Poor [15] formulated the rate-distortion in semantic communication as

$$R(D_s, D_w) = \min I(W; \hat{X}, \hat{W}) \tag{8}$$

where D_s is the semantic distortion between source, X, and recovered information, $\hat{X}$, at the receiver. D_w is the distortion between semantic representation, W, and received semantic representation, $\hat{W}$.

In recent years, the fusion of semantic communication algorithms and learning theory [16–19] is also driving the continuous development of the communication architecture. Although these explorations did not fully open the door to semantic communication, they provided us important insights to move forward theoretically.

2.3. Semantic Communication as a Lewis Signaling Game with Knowledge Bases

Jinho and Jihong [5] proposed a SC model based on the Lewis signaling game, which derived some interesting results. This work really gives us some motivations to conduct this work in some degree. Further, we make a generalization of this model with adding the asynchronous knowledge scaling up updates of the sender and the receiver, which may close to the real semantic communications in the era of intelligent communication. Thus, we briefly re-describe it.

Suppose there is a semantic communication system where Alice is the sender and Bob is the receiver. Let $T \in \mathcal{T}$ denote the semantic type that includes semantic information or messages. Alice wishes to transmit T to Bob by sending a signal $S \in \mathcal{S}$, and Bob chooses its response $R \in \mathcal{R}$. Both Alice and Bob utilize their local knowledge bases $\mathcal{K}_A$ and $\mathcal{K}_B$, respectively. The semantic architecture can be described as

$$\begin{array}{ccccc} K_A & & K_B & & \\ \downarrow & & \downarrow & & \\ T & \rightarrow & S & \rightarrow & R\ (=\hat{T}). \end{array}$$

If $R = T$, one communication process can be successful. Moreover, the reward for one communication is defined as:

$$u = \begin{cases} 1, & \text{if } R = T; \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Our object is to maximize the average reward. We use *the success rate of semantic agreement (SRSA)* to represent it.

The knowledge base can be regarded as the side information. Alice has her knowledge base $\mathcal{K}_A$. The instance of Alice's knowledge base at each time is $K_A \in \mathcal{K}_A$, which affects the generation of semantic types. On the other hand, Bob has his knowledge base $\mathcal{K}_B$ and its instance is $K_B \in \mathcal{K}_B$. K_B is the "core", used to infer the intended message together with the received signal.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0399.v1

Based on these foundations, Jinho and Jihong [5] derived some interesting results. On the one hand, the similarity of the two knowledge instances plays a crucial role in SC. With a limited bandwidth, the communication quality is regulated by $I(K_A; K_B|S)$. On the other hand, the SRSA is dependent on the similarity of the two communication parties' knowledge. However, it did not clearly setup the close relationship with Shannon's communication model. In this work, we will make a generalization of the model by adding the asynchronous knowledge scaling up updates of the sender and the receiver, which may closely reflect the quickly surging of semantic communications in many scenarios of applications.

3. Semantic Communication Models

In this section, we will present the generalized model of semantic communication from two different aspects, one is trying to setup a close relationship with Shannon communication model, and the other is trying to find the feasible modification of the model so that it can exactly reflect the quickly surging of semantic communications from the user requirements. In this regard, we first setup a basic model, and then to extend it by adding the knowledge base update functional mode with information framework expansion and knowledge collision.

3.1. Motivation

Shannon's communication system is concerned with the accurate transmission of symbols over channels. However, it does not take into account the differences in the knowledge backgrounds of two parties, which affect whether the communication can convey the desired meaning. In other words, the matching of knowledge can also be treated as another special channel, although it may not actually exist in an explicit form. In this work, we denote it as a virtual channel. We want to fill this gap and propose an intelligent communication system with coexistence of the real channel and virtual channel, making information theory more complete. Thus, we consider not only the physical channel for symbol transmission, but also the virtual channel for the transfer of knowledge between two parties. Moreover, the knowledge of both parties is also evolving, so we take into account the asynchronous knowledge scaling up updates and semantic expansion.

3.2. Basic Model

Let us first observe the *channel*, which is the physical medium of message exchange, and play the key role in the model setup of a complete semantic communication system. A *communication channel* is a system in which the output depends on probabilistically on its input [2]. For simplification, we still use Alice and Bob as the two parties participating in the communication. Suppose Alice and Bob are at opposite ends of a memoryless channel, which can transmit physical signals. The channel is said to be memoryless if the probability distribution of the output depends only on the input at that time and is conditionally independent of previous inputs or outputs. That is, the output of the channel is only related to the input at the current moment and no any feedback processing is considered here..

We use S and $\hat{S}$ to represent the input and output of the channel. Under the interference of the noise, the signal S becomes $\hat{S}$ through channel transmission. Specifically, Alice observes the input S of the channel, and Bob observes the output $\hat{S}$. Let K_A and K_B represent the knowledge instance possessed by Alice and Bob, respectively. On the one hand, Alice encodes S as T with K_A. This means that Alice uses the knowledge instance K_A to process the signal S, resulting in the semantic type T, i.e., $(S, K_A) \rightarrow T$. On the other hand, Bob decodes $\hat{S}$ into the response R with K_B, which can be expressed as $(\hat{S}, K_B \rightarrow R)$. The architecture of this process can be expressed as

$$
\begin{array}{ccccccc}
T & \leftarrow & S & \xrightarrow{(a)} & \hat{S} & \rightarrow & R\,(=\hat{T}) \\
 & & \uparrow & & \uparrow & & \\
 & & K_A & \overset{(b)}{\dashrightarrow} & K_B & &
\end{array}
\tag{10}
$$

We denote the process $(S, K_A) \rightarrow T$ as semantic encoding and $(\hat{S}, K_B) \rightarrow \hat{T}$ as semantic decoding. They indicate the understanding of meaning at the semantic level by both parties.

The definitions of symbols are summarized as:

- *Semantic Types*: $T = t_k, k = 1, \ldots, |\mathcal{T}|$, is a random variable that is generated by Alice.
- *Signals*: $S = s_l, l = 1, \ldots, |\mathcal{S}|$, is a signal that Alice sends to Bob.
- *Responses*: $R = r_n, n = 1, \ldots, |\mathcal{R}|$, is a response that Bob chooses.
- *Knowledge instances*: $K_A \in \mathcal{K}_A$ and $K_B \in \mathcal{K}_B$ represent the knowledge instance used by Alice and Bob when semantic encoding and decoding, respectively.

Under the knowledge instance K_A and K_B, the mutual information between T and $\hat{T}$ is given by

$$I(T;\hat{T}) = I(S;\hat{S}) + I(K_A;K_B|S,\hat{S}) + I(S;K_B|\hat{S}) + I(\hat{S};K_A|S) \tag{11}$$

where $I(K_A;K_B|S,\hat{S})$, $I(S;K_B|\hat{S})$ and $I(\hat{S};K_A|S)$ represent the conditional mutual information.

Proof. Since $(S, K_A) \rightarrow T$ and $(\hat{S}, K_B) \rightarrow \hat{T}$, we have $H(T) = H(K_A, S)$ and $H(\hat{T}) = H(K_B, \hat{S})$. By the characteristics of entropy and mutual information, it follows that

$$
\begin{aligned}
I(T;\hat{T}) &= H(T) - H(T|\hat{T}) && (12)\\
&= H(K_A, S) - H(K_A, S|K_B, \hat{S}) && (13)\\
&= H(S) + H(K_A|S) - H(S|K_B, \hat{S}) - H(K_A|K_B, S, \hat{S}) && (14)\\
&= H(S) - H(S|K_B, \hat{S}) + H(K_A|S) - H(K_A|K_B, S, \hat{S}) && (15)\\
&= I(S;K_B, \hat{S}) + I(K_A;K_B, \hat{S}|S) && (16)\\
&= I(S;\hat{S}) + I(K_A;K_B|S,\hat{S}) + I(S;K_B|\hat{S}) + I(\hat{S};K_A|S), && (17)
\end{aligned}
$$

which completes the proof. □

The mutual information between T and $\hat{T}$ reflects the effectiveness of communication. It indicates the highest rate in bits per channel use at which information can be sent with arbitrarily low probability of error. We note that $I(T;\hat{T})$ consists of three terms:

1. $I(S;\hat{S})$, the mutual information between the input and output of the channel. It corresponds to the *channel capacity* in Shannon's information theory.
2. $I(K_A;K_B|S,\hat{S})$, the mutual information between K_A and K_B given S and $\hat{S}$. It indicates the amount of information that two knowledge instances contain about each other when signals are known.
3. $I(S;K_B|\hat{S}) + I(\hat{S};K_A|S)$, the conditional mutual information between S and K_B, $\hat{S}$ and K_A.

From Eq. (11), we know that if Bob wants to fully understand what Alice means, the communication process needs to meet two conditions, one is the accuracy of the symbols during channel transmission, and the other is the matching degree of the knowledge bases of both parties. In other words, i) The transmission $S \rightarrow \hat{S}$ should be reliable. It indicates the effect of channel noise on the signal. Moreover, the carrier of this process actually exists, we call it an *explicit channel*. ii) The two knowledge instances should be similar.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0399.v1

Although there is no actual transmission between K_A and K_B, we assume that there is a virtual channel that reflects the probabilistic relationship between knowledge instances, which we call an *implicit channel*. The *explicit channel* and *implicit channel* together form the transmission medium of a communication system. Specifically, the characteristics of the explicit channel determine the value of the first term in Eq. (11), and the implicit channel determines the second term. In addition, the last term is affected by both explicit and implicit channels.

If the communication system has only the explicit channel, it degenerates to the Shannon case, and $I(T;\hat{T}) = I(S;\hat{S})$. Furthermore, We discuss three special cases.

I. **The explicit channel is noiseless, i.e., $S = \hat{S}$.** In this setting, the formula (11) can be simplified to

$$I(T;\hat{T}) = H(S) + I(K_A;K_B|S), \tag{18}$$

which is consistent with the conclusion in [5]. The mutual information between T and $\hat{T}$ is subjected to the entropy of the signal S and the conditional mutual information between K_A and K_B given S.

II. **The implicit channel is noiseless, i.e., $K_A = K_B$.** This means that Alice and Bob have the same knowledge instance, so the communication performance will not be affected by the difference in the background of both parties. $I(T;\hat{T})$ can be expressed as:

$$I(T;\hat{T}) = I(S;\hat{S}) + H(K_A) - I(K_A;S;\hat{S}) \tag{19a}$$
$$I(T;\hat{T}) = I(S;\hat{S}) + H(K_B) - I(K_B;S;\hat{S}), \tag{19b}$$

where $I(K_A;S;\hat{S}) = I(K_A;S) - I(K_A;S|\hat{S})$ is the mutual information between K_A, S and $\hat{S}$. If the implicit channel is noiseless, the mutual information between T and $\hat{T}$ satisfies

$$I(S;\hat{S}) \overset{(a)}{\leq} I(T;\hat{T}) \overset{(b)}{\leq} I(S;\hat{S}) + H(K_A). \tag{20}$$

Moreover, if K_A is a function of S and $\hat{S}$, the left equation (a) holds; if K_A is independent of the signal S or $\hat{S}$, the right equation (b) holds.

Proof. Since the nonnegativity of mutual information, $I(K_A;S;\hat{S}) \geq 0$. Thus, $I(T;\hat{T})$ can be upper bounded by $I(S;\hat{S}) + H(K_A)$. By the fact that the mutual information is lower than the entropy, we have $H(K_A) - I(K_A;S;\hat{S}) \geq 0$. Combining these results, we obtain

$$I(S;\hat{S}) \leq I(T;\hat{T}) \leq I(S;\hat{S}) + H(K_A), \tag{21}$$

which completes the proof of the bound of $I(T;\hat{T})$. □

III. **Both explicit and implicit channels are noiseless, i.e., $S = \hat{S}, K_A = K_B$.** In this setting, the mutual information between T and $\hat{T}$ equals the entropy of T or $\hat{T}$,

$$I(T;\hat{T}) = H(T) = H(\hat{T}). \tag{22}$$

Eq. (22) indicates that Bob can perfectly understand the meaning of Alice when channels are noiseless. That is, no information is lost.

The following result shows the bounds of SRSA, constrained by the characteristics of the explicit and implicit channels. The SRSA satisfies

$$\text{SRSA} = \Pr\{T = \hat{T}\} \leq 1 - \frac{H(K_A, S|K_B, \hat{S}) - 1}{\log|\mathcal{T}|} \tag{23}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0399.v1

Proof. Since $(K_B, \hat{S})$ can be seen as an estimator for (K_A, S). Let $P_e = \Pr\{T \neq \hat{T}\}$, then by Fano's Inequality, we obtain

$$P_e \geq \frac{H(K_A, S|K_B, \hat{S}) - 1}{\log |\mathcal{T}|} \tag{24}$$

Since SRSA$= 1 - P_e$, we can obtain Eq. (23), which completes the proof. □

3.3. EXP-SC Model

Let S_1 and S_2 denote the input signals of the explicit channel, and $\hat{S}_1$ and $\hat{S}_2$ are their corresponding output signals. Similarly, K_A^1 and K_A^2 represent Alice's knowledge instances. Alice encodes S_1 as T_1 with K_A^1, and S_2 as T_2 with K_A^2. On the other hand, Bob uses K_B^1 and K_B^2, decoding $\hat{S}_1$ and $\hat{S}_2$ into $\hat{T}_1$ and $\hat{T}_2$, respectively. In the basic model, the encoding and decoding processes of the two transmissions are

$$S_1, K_A^1 \rightarrow T_1 \quad \hat{S}_1, K_B^1 \rightarrow \hat{T}_1 \tag{25a}$$
$$S_2, K_A^2 \rightarrow T_2 \quad \hat{S}_2, K_B^2 \rightarrow \hat{T}_2 \tag{25b}$$

Expansion combines simple things into complex systems and even generates intelligence, which is consistent with the evolution of the human language system. We extend the basic model proposed in Section 2.3, based on expansion. In this context, Alice wishes to send an expansion of multiple signals to Bob. We first consider the case of two signals, which can be generalized to more. Now, what Alice sends is expanded from $S = S_1$ to $S = S_1 \oplus T_2$. We use $\oplus$ to denote semantic expansion. Specifically, the expansion of signals only represents their combination. For instance, '*Carol published a paper*' expands to '*Carol published a paper in* ***IEEE Communications Letters***'.

As known, expansion often implies collision and fusion. Similarly, the expansion of signals corresponds to the collision of knowledge bases in this work. For Alice, we use $K_A = K_A^1 \overset{\alpha}{\odot} K_A^2$ to represent the collision process, where $\odot$ denotes the collision and α is the *collision factor*. Specifically, α is between 0 and 1, determined by the task. The collision factor reflects the role of K_A^2 compared to K_A^1 when the collision occurs. It can also be understood as the relative proportion of the contribution to the newly generated knowledge instance. The process of collision represents the asynchronous knowledge scaling up updates. Similarly, we use $K_B = K_B^1 \overset{\beta}{\odot} K_B^2$ to represent the knowledge collision of Bob, where β is a collision factor. The expansion and collision process can proceed continuously as follows:

$$S = S_1 \oplus S_2 \oplus S_3 \cdots \oplus S_n \tag{26a}$$
$$K_A = K_A^1 \overset{\alpha_1}{\odot} K_A^2 \overset{\alpha_2}{\odot} K_A^3 \cdots \overset{\alpha_{n-1}}{\odot} K_A^n \tag{26b}$$
$$K_B = K_B^1 \overset{\beta}{\odot} K_B^2 \overset{\beta_2}{\odot} K_B^3 \cdots \overset{\beta_{n-1}}{\odot} K_B^n. \tag{26c}$$

Without loss of generality, the one step expansion architecture of semantic communications is described as

$$\begin{array}{ccccccc} & & S & \xrightarrow{(a)} & \hat{S} & & \\ & & \Uparrow & & \Downarrow & & \\ T & \leftarrow & S_1 \oplus S_2 & & \hat{S_1} \oplus \hat{S_2} & \rightarrow & R\ (= \hat{T}) \\ & & \uparrow \text{(c)} & & \uparrow \text{(d)} & & \\ & & K_A^1 \overset{\alpha}{\odot} K_A^2 & \overset{(b)}{\dashrightarrow} & K_B^1 \overset{\beta}{\odot} K_B^2 & & \end{array} \tag{27}$$

We named it EXP-SC. Moreover, $H(S_1 \oplus S_2, K_A^1 \overset{\alpha}{\odot} K_A^2)$ is called the Measure of Comprehension and Interpretation (MCI), which reflects the generation and evolution of semantics. It should be noted that T is not a simple logic combination of T_1 and T_2, i.e., $T \neq T_1 \oplus T_2$. For example, T_1 is *'Apple Inc.'*, it is a company and S_1 can be *'Apple'*. T_2 is *'the thirteenth generation'*, it is a number and S_2 can be *'thirteen'*. But their collision may give rise to a new word called *'iphone'*, which is a mobile communication product. In particular, T reflects the result of S under the influence of knowledge collision.

Based on these definitions above, we get some new results. The mutual information between T and $\hat{T}$ is given by

$$I(T;\hat{T}) = I(S;\hat{S}) + I(K_A^1 \overset{\alpha}{\odot} K_A^2; K_B^1 \overset{\beta}{\odot} K_B^2 | S, \hat{S}) + I(S; K_B^1 \overset{\beta}{\odot} K_B^2 | \hat{S}) + I(\hat{S}; K_A^1 \overset{\alpha}{\odot} K_A^2 | S) \tag{28}$$

Proof. It is similar to the proof of Lemma 3.2, we omit the proof. □

Eq. (28) indicates that besides the characteristics of explicit and implicit channels, the relationship between α and β also affects the performance of communication.

When the explicit channel is noiseless, the gain brought by semantic expansion is given by

$$I(T;\hat{T}) - I(T_1;\hat{T}_1) = (1+\gamma)(H(S) - H(S_1)), \tag{29}$$

where

$$\gamma = \frac{I(K_A^1 \overset{\alpha}{\odot} K_A^2; K_B^1 \overset{\beta}{\odot} K_B^2 | S_1 \oplus S_2) - I(K_A^1; K_B^1 | S_1)}{H(S_1 \oplus S_2) - H(S_1)}. \tag{30}$$

Proof. We note that

$$I(T_1;\hat{T}_1) = H(S_1) + I(K_A^1; K_B^1 | S_1). \tag{31}$$

Then,

$$\frac{I(T;\hat{T}) - I(T_1;\hat{T}_1)}{H(S) - H(S_1)} \tag{32}$$

$$= \frac{H(S) + I(K_A^1 \overset{\alpha}{\odot} K_A^2; K_B^1 \overset{\beta}{\odot} K_B^2 | S) - H(S_1) - I(K_A^1; K_B^1 | S_1)}{H(S_1 \oplus S_2) - H(S_1)} \tag{33}$$

$$= 1 + \frac{I(K_A^1 \overset{\alpha}{\odot} K_A^2; K_B^1 \overset{\beta}{\odot} K_B^2 | S) - I(K_A^1; K_B^1 | S_1)}{H(S_1 \oplus S_2) - H(S_1)}, \tag{34}$$

which completes the proof. □

Lemma 3.3 shows that the asynchronous knowledge scaling up updates determine the effect of semantic expansion.

When the explicit channel is noiseless, the mutual information between T and $\hat{T}$ is bounded by

$$\frac{1}{2}(H(S_1) + H(S_2)) \leq I(T;\hat{T}) \leq H(K_A^1 \overset{\alpha}{\odot} K_A^2) + H(S) \tag{35}$$

Proof. Since the fact that expansion would create more possibilities, leading to the increase of uncertainty, we can get

$$H(S_1 \oplus S_2) \geq H(S_1) \quad H(S_1 \oplus S_2) \geq H(S_2) \tag{36}$$

Through the properties of *non-negativity of entropy* and *conditioning reduces entropy*, Eq. (35) can be derived directly. We omit the proof. □

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0399.v1

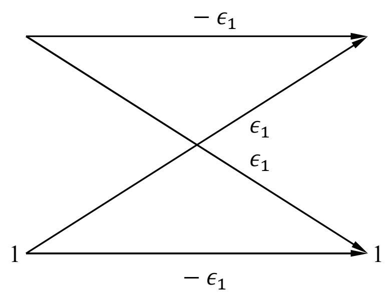


Figure 1. Binary symmetric channel.

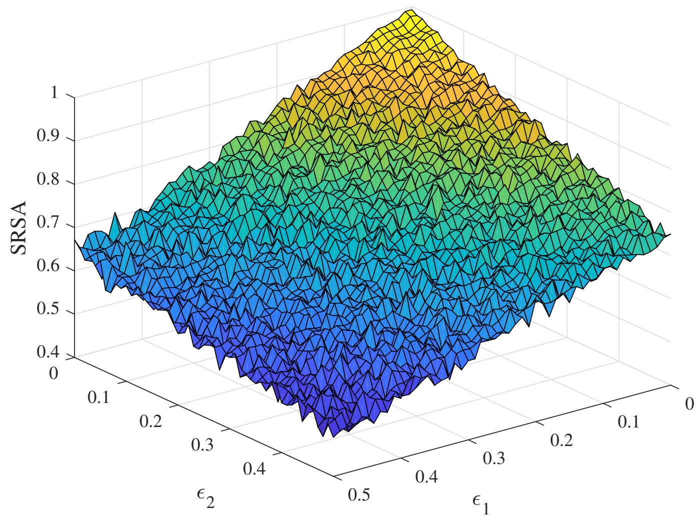


Figure 2. The simulation results of SRSA with the explicit channel error probability ϵ_1 and the implicit channel error probability ϵ_2.

4. Experiment and Numerical Results

In this section, we use SRSA to measure the performance of SC, especially the impact that asynchronous knowledge scaling up has on the system.

Basic Model. We use Q-learning in [20] to complete semantic encoding and decoding. Let $|\mathcal{S}| = M = 2$, $|\mathcal{K}_A| = |\mathcal{K}_B| = L = 2$, $|\mathcal{T}| = L^2M^2$. In addition, S, K_A and K_B are uniformly distributed. We would like to discuss the impact of the characteristics of the explicit channel and implicit channel on SC. We use the binary symmetric channel (BSC) to denote the explicit channel. The BSC is shown in Fig. 1, which shows the probabilistic relationship between S and $\hat{S}$, with the error probability ϵ_1.

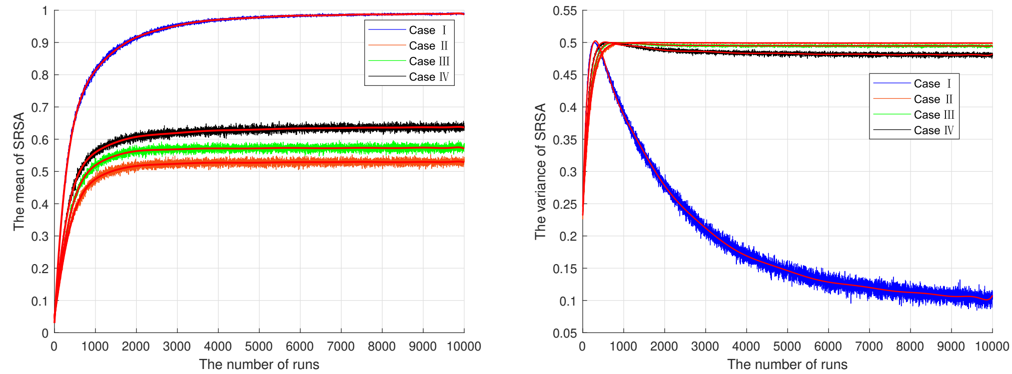
For the implicit channel, we assume the correlation between K_A and K_B satisfies that

$$K_B = \begin{cases} K_A, & \text{with probability } 1 - \epsilon_2 \\ U, & \text{with probability } \epsilon_2, \end{cases}$$

where $U \sim \text{Unif}\{1, L\}$ is an independent random variable. In Fig. 2, we show the SRSA with $\epsilon_1 \in [0, 0.5]$ and $\epsilon_2 \in [0, 0.5]$. When $\epsilon_1 = \epsilon_2 = 0$, SRSA reaches the maximum 1. As ϵ_1 and ϵ_2 increase, SASA will decrease in both directions. This indicates that two channels jointly determine the quality of communication. For high quality SC, It is necessary to meet the condition that $\hat{S} = S$ and $K_B = K_A$ with a high probability.

EXP-SC Model. We want to explore how communication quality varies in the context of different relationships between the receiver and the sender's knowledge instance. For simplicity, we assume that the explicit channel is noiseless, so we can focus on knowledge

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0399.v1



(a) The mean of SRSA (b) The variance of SRSA

Figure 3. The simulation results of SRSA with the number of runs. We divide it into four cases. Case I: $K_B^1 = K_A^1$, $K_B^2 = K_A^2$ and $\beta = \alpha$; II: $K_B^1 \neq K_A^1$ (with error probability 0.5), $K_B^2 = K_A^2$ and $\beta = \alpha$; III: $K_B^1 = K_A^1$, $K_B^2 \neq K_A^2$ (with error probability 0.5) and $\beta = \alpha$; IV: $K_B^1 = K_A^1$, $K_B^2 = K_A^2$ and $\beta = \frac{1}{2}\alpha$. (a) the mean of SRSA. (b) the variance of SRSA.

updates. That is, SRSA varies with the relationship of K_B^1, K_B^2, β and K_A^1, K_A^2, α. We categorize it into four cases.

- Case I: $K_B^1 = K_A^1$, $K_B^2 = K_A^2$ and $\beta = \alpha$. The implicit channel is noiseless. Bob has the same asynchronous knowledge scaling up updates mode as Alice. That is, the receiver has all the knowledge of the sender.
- Case II: $K_B^1 \neq K_A^1$ (with error probability 0.5), $K_B^2 = K_A^2$ and $\beta = \alpha$. The receiver has partial knowledge of the sender.
- Case III: $K_B^1 = K_A^1$, $K_B^2 \neq K_A^2$ (with error probability 0.5) and $\beta = \alpha$. The receiver has partial knowledge of the sender.
- Case IV: $K_B^1 = K_A^1$, $K_B^2 = K_A^2$ and $\beta = \frac{1}{2}\alpha$. The collision factors do not equal.

Fig. 3 illustrates the simulation results of SRSA with the number of runs, where Fig. 3 (a) is the mean and (b) is the variance. In all four cases, the mean of SRSA gradually converges to a stable region. Case I is close to 1, which implies that when the receiver has all the knowledge of the sender, it can express exactly the same semantics as the sender. In other words, Bob has the same asynchronous knowledge scaling up updates mode as Alice, which facilitates the success of SC. The curve of case II and III are almost the same, but the value of case II is always higher than that of case III. This is as expected because K_A^1 and K_B^1 play a leading role in the collision compared to K_A^2 and K_B^2. Case IV reflects the learning of SRSA when Bob and Alice have different collision factors. The value of case IV is higher than that of II and III, which indicates that the collision factor plays a smaller role than the knowledge itself in SC. These results also show that learning can improve SC quality with only partial background knowledge, but there is a limit. On the other hand, the variance of case I keeps decreasing, and the other three cases also stabilize after peaking quickly. This further suggest that learning can evolve continuously with full knowledge, but there is an upper limit to learning with partial or no knowledge.

5. Conclusion

In this paper, we presented the concept of semantic expansion and knowledge collision in SC. It represents the combination and superposition of information by the sender. Based on semantic expansion, we further proposed a semantic communication system called EXP-SC. Moreover, the semantic expansion corresponds to knowledge collision, which provides possibility for the evolution and upgrading of communication systems. On the other hand, we reached some conclusions for semantic information theory in the context of asynchronous knowledge scaling up updates, getting some bounds for SC. Specifically, the receiver's understanding of the knowledge collision and updates determines the effect of communication.

Semantic communication is evolving towards intelligence. The insights gained from this work may be of assistance to the future semantic communication system and 6G. It is also expected to pave the way for the design of next-generation real-time data networking and will provide the foundational technology for a plethora of socially useful services, including autonomous transportation, consumer robotics, VR/AR and metaverse.

Author Contributions: Conceptualization, P.F. and G.X.; methodology, P.F. and G.X.; software, G.X.; formal analysis, P.F. and G.X.; writing—original draft preparation, G.X.; writing—review and editing, P.F. and G.X.; visualization, P.F. and G.X.; supervision, P.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (Grant NO.2021YFA1000504).

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable

Acknowledgments: We thank Prof. Jinho for the help and source code in [5].

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A Proof of the mutual information for three special channel cases

I. **The explicit channel is noiseless.** Since $S = \hat{S}$, we have $I(S;\hat{S}) = H(S)$ and $I(K_A;K_B|S,\hat{S}) = I(K_A;K_B|S)$. Moreover, S is a function of $\hat{S}$, it follow that $I(S;K_B|\hat{S}) = 0$. Similarly, $\hat{S}$ is also a function of S, we obtain $I(\hat{S};K_A|S) = 0$. Thus, the formula (11) can be simplified to

$$I(T;\hat{T}) = H(S) + I(K_A;K_B|S), \tag{A1}$$

II. **The implicit channel is noiseless**. Since $K_A = K_B$, the mutual information between T and $\hat{T}$ is written

$$\begin{aligned}
I(T;\hat{T}) &= I(S;\hat{S}) + I(K_A;K_B|S,\hat{S}) + I(S;K_B|\hat{S}) + I(\hat{S};K_A|S) &\text{(A2)}\\
&= I(S;\hat{S}) + H(K_A|S,\hat{S}) + I(S;K_A|\hat{S}) + I(\hat{S};K_A|S) &\text{(A3)}\\
&= I(S;\hat{S}) + H(K_A) - I(K_A;S,\hat{S}) + I(S;K_A|\hat{S}) + I(\hat{S};K_A|S) &\text{(A4)}\\
&= I(S;\hat{S}) + H(K_A) - I(K_A;S) - I(K_A;\hat{S}|S) + I(S;K_A|\hat{S}) + I(\hat{S};K_A|S) &\text{(A5)}\\
&= I(S;\hat{S}) + H(K_A) - I(K_A;S) + I(K_A;S|\hat{S}) &\text{(A6)}\\
&= I(S;\hat{S}) + H(K_A) - I(K_A;S;\hat{S}). &\text{(A7)}
\end{aligned}$$

It can also be expressed as

$$I(T;\hat{T}) = I(S;\hat{S}) + H(K_B) - I(K_B;S;\hat{S}). \tag{A8}$$

III. **Both explicit and implicit channels are noiseless, i.e., $S = \hat{S}, K_A = K_B$.** In this setting, we have $I(K_A;K_B|S) = H(K_A|S)$. Thus, Eq. (18) is written

$$I(T;\hat{T}) = H(S) + H(K_A|S) = H(K_A,S) = H(T). \tag{A9}$$

Similarly, we also obtain $I(T;\hat{T}) = H(\hat{T})$.

References

1. Shannon, C.E. A mathematical theory of communication. *The Bell system technical journal* **1948**, *27*, 379–423.
2. Cover, T.M.; Thomas, J.A. *Elements of information theory*; 1991.

3. Shi, G.; Xiao, Y.; Li, Y.; Xie, X. From semantic communication to semantic-aware networking: Model, architecture, and open problems. *IEEE Communications Magazine* **2021**, *59*, 44–50.
4. Weaver, W. Recent contributions to the mathematical theory of communication. *ETC: a review of general semantics* **1953**, pp. 261–281.
5. Choi, J.; Park, J. Semantic Communication as a Signaling Game with Correlated Knowledge Bases. *arXiv preprint arXiv:2209.00280* **2022**.
6. Carnap, R.; Bar-Hillel, Y.; et al. An outline of a theory of semantic information **1952**.
7. Bao, J.; Basu, P.; Dean, M.; Partridge, C.; Swami, A.; Leland, W.; Hendler, J.A. Towards a theory of semantic communication. In Proceedings of the 2011 IEEE Network Science Workshop. IEEE, 2011, pp. 110–117.
8. Floridi, L. Outline of a theory of strongly semantic information. *Minds and machines* **2004**, *14*, 197–221.
9. Qin, Z.; Tao, X.; Lu, J.; Li, G.Y. Semantic communications: Principles and challenges. *arXiv preprint arXiv:2201.01389* **2021**.
10. Shi, G.; Gao, D.; Song, X.; Chai, J.; Yang, M.; Xie, X.; Li, L.; Li, X. A new communication paradigm: from bit accuracy to semantic fidelity. *arXiv preprint arXiv:2101.12649* **2021**.
11. D'Alfonso, S. On quantifying semantic information. *Information* **2011**, *2*, 61–101.
12. Kolchinsky, A.; Wolpert, D.H. Semantic information, autonomous agency and non-equilibrium statistical physics. *Interface focus* **2018**, *8*, 20180041.
13. Kountouris, M.; Pappas, N. Semantics-empowered communication for networked intelligent systems. *IEEE Communications Magazine* **2021**, *59*, 96–102.
14. Tucker, M.; Shah, J.; Levy, R.; Zaslavsky, N. Towards Human-Agent Communication via the Information Bottleneck Principle. *arXiv preprint arXiv:2207.00088* **2022**.
15. Liu, J.; Zhang, W.; Poor, H.V. A rate-distortion framework for characterizing semantic information. In Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT). IEEE, 2021, pp. 2894–2899.
16. Weng, Z.; Qin, Z. Semantic communication systems for speech transmission. *IEEE Journal on Selected Areas in Communications* **2021**, *39*, 2434–2444.
17. Huang, D.; Tao, X.; Gao, F.; Lu, J. Deep learning-based image semantic coding for semantic communications. In Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM). IEEE, 2021, pp. 1–6.
18. Zhou, Q.; Li, R.; Zhao, Z.; Peng, C.; Zhang, H. Semantic communication with adaptive universal transformer. *IEEE Wireless Communications Letters* **2021**, *11*, 453–457.
19. Zhang, H.; Shao, S.; Tao, M.; Bi, X.; Letaief, K.B. Deep Learning-Enabled Semantic Communication Systems with Task-Unaware Transmitter and Dynamic Data. *arXiv preprint arXiv:2205.00271* **2022**.
20. Catteeuw, D.; Manderick, B. The limits and robustness of reinforcement learning in Lewis signalling games. *Connection Science* **2014**, *26*, 161–177.