The Semantic Communication Game

Basak Guler
The Pennsylvania State University
University Park, PA
basak@psu.edu

Aylin Yener
The Pennsylvania State University
University Park, PA
yener@ee.psu.edu

Ananthram Swami
Army Research Laboratory
Adelphi, MD
a.swami@ieee.org

***Abstract*—We study how to communicate semantic information in the presence of an agent that can influence the decoder by providing side information. The agent's true intentions, which may be adversarial or helpful, is unknown to the communicating parties. Actions taken by the agent are governed by its intentions, and they may improve or deteriorate the communication performance. We characterize the optimal transmission policies to minimize the end-to-end average semantic error, i.e., difference between the meanings of intended and recovered messages, under the uncertainty in the agent's true intentions. We formulate the semantic communication problem as a Bayesian game, and investigate the conditions under which a pure strategy Bayesian Nash equilibrium exists. We then explore the structure of the encoding and decoding functions under the mixed strategy Bayesian Nash equilibrium, which for the semantic communication problem at hand always exists. Our results show that the optimal policies are strongly influenced by the belief the parties hold about the agent's true intention.**

I. Introduction

Communication in the real world is often influenced by the external information available to the interacting parties. Humans, for instance, are influenced by their communities, social media, or news sources while interpreting the events that occur around them to form their opinions. Not all information, however, is useful, in that some sources provide a better understanding of the world, while others can be misleading.

Game theoretic formulations have provided a comprehensive framework for modeling the interplay between the actors of a networked system [1]. Bayesian games for example, consider games with incomplete information due to an uncertainty about the characteristics of one or more players. The objective of each player, i.e., its payoff function, is determined by its characteristics. Bayesian games have diverse applications including network intrusion detection [2], wireless spectrum utilization [3], and power allocation [4].

The performance criterion in modern communication systems is based on error rates that do not take into account the semantics of communicated messages. In such systems, error rates between semantically similar words, such as *car* and *automobile*, are treated equally as semantically distant words, such as *car* and *computer*. On the other hand, while communicating a message *car*, its meaning would be preserved much better if an *automobile* is recovered and not a *computer*. Semantic similarity quantifies the distance between the meanings of two words [5], [6], with applications ranging from artificial intelligence to natural language processing and information retrieval. These measures are often based on a thesaurus, e.g., WordNet [7], or statistics from a large corpus. An index assignment scheme is developed in [8] for maximizing the semantic similarity between intended and recovered messages in a noisy channel.

This research is sponsored by the U.S. Army Research Laboratory under the Network Science Collaborative Technology Alliance, Agreement Number W911NF-09-2-0053.

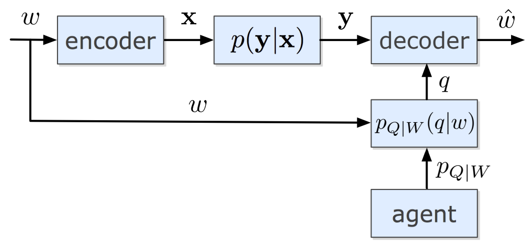

Fig. 1. Semantic communication channel.

This paper studies how to communicate semantic information in the presence of an influential entity. This individual, which we refer to as an *agent* in the sequel, can influence the decoder by providing side information. Its true nature, whether adversarial or helpful, is unknown to the communicating parties. We envision that an agent with such influence capability may have a significant impact on information transfer. For instance, an adversarial agent would target causing errors in message recovery, while a helpful agent would act to facilitate effective communication. It is therefore essential to tailor the transmission policies to take into account the uncertainty in the intentions of influential entities.

We view the semantic communication problem with external influence as a Bayesian game played between the encoder/decoder and the agent. The encoder/decoder pair wishes to minimize the average semantic error in the recovered messages. Depending on its characteristic, the agent may aim at helping or harming the communication performance. We investigate the conditions for which a pure strategy Bayesian Nash equilibrium exists, and characterize the mixed strategy Bayesian Nash equilibrium. In our numerical studies, we implement the semantic communication game to determine the equilibrium strategies as well as the structure of the semantic error minimizing encoding and decoding functions. Our results show that judicious transmission policies can greatly reduce the errors that occur in the meanings of recovered words.

In the remainder of the paper, we denote x for a scalar, $\mathbf{x}$ for a vector, $\mathcal{X}$ for a set with cardinality $|\mathcal{X}|$, $\mathcal{X}^n$ for the n-fold Cartesian product of $\mathcal{X}$, and $\mathbb{E}[\cdot]$ for the expectation.

II. System Model

We consider the communication scenario in Fig. 1. The encoder observes a word w from a distribution $p_W(w)$ over a set $\mathcal{W}$. To characterize the actions the agent can take to influence the decoder, we define a set $\mathcal{P}$. Each element of $\mathcal{P}$ is a random variable $p_{Q|W}$ such that for each $w \in \mathcal{W}$, $p_{Q|W} \in \mathcal{P}$ represents a probability distribution over a set of contexts $\mathcal{Q}$ such that $\sum_{q\in\mathcal{Q}} p_{Q|W}(q|w) = 1$. The agent chooses $p_{Q|W}$ in accordance with its characteristic, e.g., friend or foe. Sets $\mathcal{W}$, $\mathcal{P}$, and $\mathcal{Q}$ have finite cardinality. The decoder then observes a context $q \sim p_{Q|W}(q|w)$. In essence, $p_{Q|W}$ represents the *external information* provided by the agent to the decoder.

The encoder wishes to transmit the word w in a way that the *meaning* of the word recovered by the decoder is close to that of w. We focus on a deterministic encoding function given by $g: \mathcal{W} \rightarrow \mathcal{X}^{(n)}$, which maps each w to a vector $\mathbf{x} = g(w)$ of length $n \in \mathbb{Z}^+$. Each element of $\mathbf{x}$ is selected from a finite alphabet $\mathcal{X}$, and hence $\mathcal{X}^{(n)} \subseteq \mathcal{X}^n$. A noisy channel exists between the encoder and the decoder, characterized by the distribution $p(\mathbf{y}|\mathbf{x})$, where the channel output $\mathbf{y}$ is a vector of length n from a finite set $\mathcal{Y}^{(n)} \subseteq \mathcal{Y}^n$, such that each element of $\mathbf{y}$ is from a finite alphabet $\mathcal{Y}$. The decoder recovers a word $\hat{w} \in \mathcal{W}$ from the received $\mathbf{y}$ and context q using a decoding function $h: \mathcal{Y}^{(n)} \times \mathcal{Q} \rightarrow \mathcal{W}$. The channel output, when conditioned on the channel input, is independent from the word and the context, i.e., we have that

$$p(\mathbf{y}|\mathbf{x}, w, q) = p(\mathbf{y}|\mathbf{x}) \tag{1}$$

where $\mathbf{x} = g(w)$. As a result,

$$\begin{aligned} p(w, q, \mathbf{y}) &= \sum_{\mathbf{x}'\in\mathcal{X}^{(n)}} p(w, q, \mathbf{y}, \mathbf{x}') && (2)\\ &= \sum_{\mathbf{x}'\in\mathcal{X}^{(n)}} p(\mathbf{y}|\mathbf{x}', w, q) p(\mathbf{x}'|w, q) p_{Q|W}(q|w) p_W(w) && (3)\\ &= p(\mathbf{y}|\mathbf{x}, w, q) p_{Q|W}(q|w) p_W(w) && (4)\\ &= p(\mathbf{y}|\mathbf{x}) p_{Q|W}(q|w) p_W(w) && (5) \end{aligned}$$

where $\mathbf{x} = g(w)$, (3) is from the chain rule of probability, (4) holds since $p(\mathbf{x}'|w, q) = 1$ if and only if $\mathbf{x}' = g(w) = \mathbf{x}$ for a deterministic encoder, and (5) is from (1).

We define the semantic distance between two words as

$$d(w, \hat{w}) = 1 - sim(w, \hat{w}), \quad w, \hat{w} \in \mathcal{W}, \tag{6}$$

where $0 \leq sim(w, \hat{w}) \leq 1$ denotes the semantic similarity between w and $\hat{w}$. The expected semantic error for some g, h, and $p_{Q|W} \in \mathcal{P}$ is

$$\begin{aligned} D(g, h, p) &= \sum_{\substack{w\in\mathcal{W}, q\in\mathcal{Q}, \mathbf{y}\in\mathcal{Y}^{(n)}:\\ h(\mathbf{y},q)\neq w}} p(w, q, \mathbf{y}) d(w, h(\mathbf{y}, q)) && (7)\\ &= \sum_{\substack{w\in\mathcal{W}, q\in\mathcal{Q}, \mathbf{y}\in\mathcal{Y}^{(n)}:\\ h(\mathbf{y},q)\neq w, \mathbf{x}=g(w)}} p(\mathbf{y}|\mathbf{x}) p_{Q|W}(q|w) p_W(w) d(w, h(\mathbf{y}, q)) && (8) \end{aligned}$$

since whenever $h(\mathbf{y}, q) = w$, then $d(w, h(\mathbf{y}, q)) = 0$, and (8) follows from (5). In the sequel, for notational simplicity, we let $p_W(w) = p(w)$ and $p_{Q|W} = p$ with $p_{Q|W}(q|w) = p(q|w)$.

The agent may have an adversarial or helpful characteristic, which is unknown to the encoder decoder pair. The uncertainty about the agent's true characteristic is termed in game theory literature [9]; as its *type*. We adopt the same terminology in the sequel. The distribution of types $\Theta = \{a, h\}$ is given by

$$p(\theta) = \begin{cases} \alpha & \text{if} \quad \theta = a \\ 1 - \alpha & \qquad \theta = h \end{cases} \tag{9}$$

such that $0 \leq \alpha \leq 1$ where $\theta = a$ and $\theta = h$ denote whether the agent is adversarial or helpful, respectively. That is, $p(\theta)$ represents the encoder's belief on the agent's true intentions.

III. Bayesian Game Formulation

We formally define the semantic communication game as follows. Consider a two-player strategic Bayesian game $\langle N, (S_i), \theta, (u_i(S, \theta)), p(\theta)\rangle$. The set of players is given by $N = \{1, 2\}$. The agent is denoted by player 1. Player 2 designs the encoding and decoding protocols. S_i and $u_i(S, \theta)$ stand for the strategy set and payoff (utility) of player i, respectively, where S is the set of strategy profiles.

The strategy set of player 1 is $S_1 = \mathcal{P}$, since it can pick any element of $\mathcal{P}$ as a pure strategy. The set of pure strategies player 2 can follow consists of all pairs of mappings (g, h) such that $g: \mathcal{W} \rightarrow \mathcal{X}^{(n)}$ and $h: \mathcal{Y}^{(n)} \times \mathcal{Q} \rightarrow \mathcal{W}$. The strategy set of player 2 is then defined as $S_2 = \mathcal{G} \times \mathcal{H}$, where $\mathcal{G}$ and $\mathcal{H}$ are the set of all valid g and h functions, respectively.

Player 1 belongs to one of the following types: i) adversary, denoted by $\theta = a$, ii) helpful, denoted by $\theta = h$. The exact type of player 1 is private information that is available only to the agent, and is unknown to player 2. We utilize the definition of $p(\theta)$ from (9) to represent the uncertainty about the type of player 1. Player 2 has a single type, whose goal is to always minimize (8), and this is known to both sides. Hence, without loss of generality, we only demonstrate the types of player 1.

If player 1 is of adversarial type, it wishes to maximize the average semantic error given in (8), whereas if it is of helpful type, it wishes to minimize it. Accordingly, we define the payoff function for player 1 as

$$u_1(p_\theta, (g, h), \theta) = \begin{cases} D(g, h, p_\theta) & \text{if} \quad \theta = a \\ -D(g, h, p_\theta) & \qquad \text{o.w.} \end{cases} \tag{10}$$

and the payoff function for player 2 as

$$u_2((g, h), p_\theta, \theta) = -D(g, h, p_\theta), \tag{11}$$

given $(g, h) \in S_2$, $p_\theta \in S_1$, and $\theta \in \Theta$, where we denote p_θ as the strategy taken by player 1 with type θ. The two players choose their strategies independently before the communication takes place. We consider a non-cooperative strategic game in which the decision of each player is independent of the decision of the other player. Both players are rational.

IV. Pure Strategy Bayesian Nash Equilibrium

We study in this section the conditions under which a pure strategy Bayesian Nash Equilibrium exists.

A. Almost Helpful Agent and an Optimistic Strategy

The following conditions identify the optimal strategies of player 2 against an *almost helpful* player 1. They state that, if player 1 is strongly believed to be helpful, then an *optimistic* encoding and decoding strategy, one that is designed for a strictly helpful agent, creates a Bayesian Nash equilibrium.

Theorem 1. *Let*

$$p_a^* = \arg\max_{p_a \in \mathcal{P}} D(g^*, h^*, p_a), \tag{12}$$

$$p_h^* = \arg\min_{p_h \in \mathcal{P}} D(g^*, h^*, p_h), \tag{13}$$

$$(g^*, h^*) = \arg\min_{(g,h)\in\mathcal{G}\times\mathcal{H}} D(g, h, p_h^*), \tag{14}$$

such that (g^*, h^*) *is the strategy of player* 2*, and* p_θ^* *is the strategy taken by player* 1 *if it is of type* $\theta \in \{a, h\}$*. Define*

$$\beta(g,h) = D(g,h,p_h^*) - D(g^*,h^*,p_h^*), \tag{15}$$

$$\mu(g,h) = D(g^*,h^*,p_a^*) - D(g,h,p_a^*) + D(g,h,p_h^*) - D(g^*,h^*,p_h^*). \tag{16}$$

Then, $(p_a^*, p_h^*, (g^*, h^*))$ *is a pure strategy Bayesian Nash equilibrium if*

i) $$\alpha \le \min_{(g,h)\in\mathcal{A}} \frac{\beta(g,h)}{\mu(g,h)}, \tag{17}$$

ii) if $\mathcal{B} \neq \emptyset$*, then* $\alpha = 0$ *or* $\mu(g,h) \le 0 \;\; \forall (g,h) \in \mathcal{B}$, (18)

where

$$\mathcal{A} = \{(g,h) \in \mathcal{G}\times\mathcal{H} : \beta(g,h) > 0, \mu(g,h) > 0, (g,h) \neq (g^*,h^*)\} \tag{19}$$

$$\mathcal{B} = \{(g,h) \in \mathcal{G}\times\mathcal{H} : \beta(g,h) = 0, (g,h) \neq (g^*,h^*)\}. \tag{20}$$

Proof. We show that whenever (17) holds, at $(p_a^*, p_h^*, (g^*, h^*))$, both players play their best responses to one another. If player 1 is of adversarial type, then

$$\begin{aligned} u_1(p_a^*, (g^*,h^*), a) &= D(g^*,h^*,p_a^*) && (21)\\ &= \max_{p_a\in\mathcal{P}} D(g^*,h^*,p_a) && (22)\\ &\ge D(g^*,h^*,p_a) && (23)\\ &= u_1(p_a, (g^*,h^*), a) && (24) \end{aligned}$$

for all $p_a \neq p_a^*$ such that $p_a \in \mathcal{P}$, where (21) is from (10). It follows from (24) that p_a^* is player 1's best response to player 2's strategy (g^*, h^*), if player 1 is adversarial. If instead player 1 is of helpful type,

$$\begin{aligned} u_1(p_h^*, (g^*,h^*), h) &= -D(g^*,h^*,p_h^*) && (25)\\ &= -\min_{p_h\in\mathcal{P}} D(g^*,h^*,p_h) && (26)\\ &\ge -D(g^*,h^*,p_h) && (27)\\ &= u_1(p_h, (g^*,h^*), h) && (28) \end{aligned}$$

for all $p_h \neq p_h^*$ such that $p_h \in \mathcal{P}$, where (25) is from (10). From (28) we observe that p_h^* is player 1's best response to player 2's strategy (g^*, h^*), if player 1 is helpful. Therefore, whenever player 2 adopts the strategy (g^*, h^*), player 1's best response is p_a^* if it is adversarial, and p_h^* if it is helpful.

To find the best response of player 2, we determine its expected payoff when player 1 takes the strategy p_a^* if it is adversarial and p_h^* if it is helpful. Player 2's expected payoff for an arbitrary strategy $(g,h) \in \mathcal{G}\times\mathcal{H}$ is

$$\begin{aligned} &\mathbb{E}[u_2((g,h), p_\theta, \theta)] \\ &\quad = \alpha u_2((g,h), p_a^*, a) + (1-\alpha) u_2((g,h), p_h^*, h) && (29)\\ &\quad = -\left(\alpha D(g,h,p_a^*) + (1-\alpha) D(g,h,p_h^*)\right). && (30) \end{aligned}$$

To prove that (g^*, h^*) is player 2's best response, one that maximizes its expected payoff in (30), it remains to show that

$$\mathbb{E}[u_2((g^*,h^*), p_\theta, \theta)] \ge \mathbb{E}[u_2((g,h), p_\theta, \theta)] \tag{31}$$

for all $(g,h) \in \mathcal{G}\times\mathcal{H}$ such that $(g,h) \neq (g^*,h^*)$. We obtain

$$\begin{aligned} &\mathbb{E}[u_2((g^*,h^*), p_\theta, \theta)] - \mathbb{E}[u_2((g,h), p_\theta, \theta)] \\ &\quad = -\left(\alpha D(g^*,h^*,p_a^*) + (1-\alpha) D(g^*,h^*,p_h^*)\right) \\ &\qquad + \left(\alpha D(g,h,p_a^*) + (1-\alpha) D(g,h,p_h^*)\right) && (32)\\ &\quad = \beta(g,h) - \alpha\mu(g,h) && (33) \end{aligned}$$

from (15), (16), and (30). For $(g,h) \in \mathcal{A}$, using (16)-(17),

$$\beta(g,h) \ge \alpha\mu(g,h), \tag{34}$$

since $\mu(g,h) > 0$. By comparing (34) with (33), we have that (31) holds for all $(g,h) \in \mathcal{A}$. For $(g,h) \in \mathcal{B}$, we observe that (31) immediately holds by combining (18), (20) and (33).

Lastly, we show that (g^*, h^*) remains to be the best response for $(g,h) \notin \mathcal{A}\cup\mathcal{B}$. From the definition of (g^*, h^*) in (14),

$$D(g^*,h^*,p_h^*) \le D(g,h,p_h^*) \tag{35}$$

and therefore $\beta(g,h) \ge 0$ for all $(g,h) \in \mathcal{G}\times\mathcal{H}$. As a result, we need to investigate only the following cases

$$i) \quad \beta(g,h) > 0, \quad \mu(g,h) = 0, \tag{36}$$

$$ii) \quad \beta(g,h) > 0, \quad \mu(g,h) < 0, \tag{37}$$

for $(g,h) \notin \mathcal{A}\cup\mathcal{B}$. From (33) and since $\alpha \ge 0$, it follows by inspection that (31) is satisfied for both (36) and (37).

Therefore, (g^*, h^*) is player 2's best response to player 1's strategies p_a^* and p_h^*, where player 1 takes p_a^* if adversarial and p_h^* if helpful. At the same time, p_a^* and p_h^* are player 1's best response to player 2's strategy (g^*, h^*), for when it's adversarial or helpful, respectively. Hence, $(p_a^*, p_h^*, (g^*, h^*))$ in (12)-(14) is a pure strategy Bayesian Nash equilibrium. □

B. Almost Adversarial Agent and a Pessimistic Strategy

In this section, we study the optimal strategies for player 2 against an *almost adversary* player 1. We show that if player 1 is strongly believed to be adversarial, then a *pessimistic* encoding and decoding strategy that is designed for a strictly adversarial agent creates a Bayesian Nash equilibrium.

Theorem 2. *Define*

$$p_a^* = \arg\max_{p_a \in \mathcal{P}} D(g^*, h^*, p_a), \tag{38}$$

$$p_h^* = \arg\min_{p_h \in \mathcal{P}} D(g^*, h^*, p_h), \tag{39}$$

$$(g^*, h^*) = \arg\min_{(g,h)\in\mathcal{G}\times\mathcal{H}} D(g, h, p_a^*), \tag{40}$$

such that (g^*, h^*) *is the strategy of player* 2*, and* p_θ^* *is the strategy of player* 1 *with type* $\theta \in \{a, h\}$*. If*

$$\alpha \ge \max_{(g,h)\in\mathcal{M}} \frac{\beta(g,h)}{\mu(g,h)}, \tag{41}$$

then, $(p_a^*, p_h^*, (g^*, h^*))$ *is a pure strategy Bayesian Nash equilibrium where*

$$\mathcal{M} = \{(g,h) \in \mathcal{G}\times\mathcal{H} : \beta(g,h) \neq 0, \mu(g,h) < 0, (g,h) \neq (g^*,h^*)\} \tag{42}$$

with $\beta(g,h)$ *is defined as in* (15)*, and* $\mu(g,h)$ *as in* (16)*.*

Proof. Player 1's best response to player 2's strategy (g^*, h^*) is p_a^* if player 1 is adversarial, and p_h^* if player 1 is helpful.

The proof of this part follows the same lines as in (24) and (28) of Theorem 1, and is omitted for space considerations.

We next determine player 2's best response to maximize its expected payoff when player 1 takes the strategy p_a^* if it is adversarial and p_h^* if it is helpful. Player 2's expected payoff for a strategy $(g,h) \in \mathcal{G} \times \mathcal{H}$ is as given in (30). We show that (g^*, h^*) from (40) is player 2's best response, i.e.,

$$\mathbb{E}[u_2((g^*, h^*), p_\theta, \theta)] \geq \mathbb{E}[u_2((g,h), p_\theta, \theta)] \tag{43}$$

for all $(g,h) \in \mathcal{G} \times \mathcal{H}$ such that $(g,h) \neq (g^*, h^*)$. Following along the lines of (33), we can find that

$$\mathbb{E}[u_2((g^*, h^*), p_\theta, \theta)] - \mathbb{E}[u_2((g,h), p_\theta, \theta)] = \beta(g,h) - \alpha\mu(g,h). \tag{44}$$

For $(g,h) \in \mathcal{M}$, one has $\mu(g,h) < 0$, thus (41) is equal to

$$\alpha\mu(g,h) \leq \beta(g,h), \tag{45}$$

from which, by comparing with (44), we have that (43) holds.

Next, consider the strategies $(g,h) \notin \mathcal{M}$. Let

$$\lambda(g,h) = D(g,h,p_a^*) - D(g^*, h^*, p_a^*). \tag{46}$$

From (40),

$$\lambda(g,h) \geq 0, \quad \forall (g,h) \in \mathcal{G} \times \mathcal{H}. \tag{47}$$

Combining (15) and (16) with (46),

$$\mu(g,h) = \beta(g,h) - \lambda(g,h), \tag{48}$$

and from (47),

$$\mu(g,h) \leq \beta(g,h). \tag{49}$$

As a result of (49), we only need to inspect the following cases for a given $(g,h) \notin \mathcal{M}$: i) $\beta(g,h) > 0$, $\mu(g,h) > 0$, ii) $\beta(g,h) > 0$, $\mu(g,h) = 0$, iii) $\beta(g,h) = 0$, $\mu(g,h) < 0$, iv) $\beta(g,h) = 0$, $\mu(g,h) = 0$. We observe that (43) is immediately satisfied for ii), iii) and iv), by combining (44) with the fact that $\alpha \geq 0$. For i), (43) follows from (44) combined with

$$\alpha\mu(g,h) \leq \mu(g,h) \leq \beta(g,h), \tag{50}$$

which is a result of (49) and $0 \leq \alpha \leq 1$.

Therefore $(p_a^*, p_h^*, (g^*, h^*))$ from (38)-(40) is a pure strategy Bayesian Nash equilibrium since (g^*, h^*) is player 2's best response to player 1's strategies p_a^* and p_h^*, simultaneously, p_a^* and p_h^* are player 1's best response to (g^*, h^*). □

Remark 1. *It follows from Theorems 1 and 2 that if the agent is believed to be helpful (adversarial), then the encoder and decoder designed for a strictly helpful (adversarial) agent form a pure strategy Bayesian Nash equilibrium.*

C. Semantic Error Minimizing Encoder and Decoder

We now characterize the average semantic error minimizing encoding and decoding function for a fixed conditional distribution between the words and the contexts, i.e., $\mathcal{P} = \{p\}$, where $p = p(q|w)$. To find the optimal decoder for a fixed encoder g, we find $h \in \mathcal{H}$ that minimizes

$$D(g,h,p) = \sum_{q\in\mathcal{Q}} p(q) \sum_{\mathbf{y}\in\mathcal{Y}^{(n)}} \sum_{\substack{w\in\mathcal{W}: \\ h(\mathbf{y},q)\neq w, \mathbf{x}=g(w)}} p(w|q) p(\mathbf{y}|\mathbf{x}) d(w, h(\mathbf{y}, q)), \tag{51}$$

where $p(q) = \sum_{w\in\mathcal{W}} p(q|w)p(w)$ and $p(w|q) = \frac{p(q|w)p(w)}{p(q)}$. From (51), the optimal decoding rule can be determined as follows. For each $q \in \mathcal{Q}$, assign $h(\mathbf{y}, q) = \hat{w}$ for $\mathbf{y} \in \mathcal{Y}^{(n)}$ if

$$f_d(q, \mathbf{y}, \hat{w}) \leq f_d(q, \mathbf{y}, w') \tag{52}$$

for all $w' \in \mathcal{W}$, where

$$f_d(q, \mathbf{y}, w') = \sum_{w\in\mathcal{W}: w'\neq w, \mathbf{x}=g(w)} p(w|q) p(\mathbf{y}|\mathbf{x}) d(w, w'). \tag{53}$$

To find the optimal encoder for a fixed decoder h, we find the $g \in \mathcal{G}$ that minimizes

$$D(g,h,p) = \sum_{w\in\mathcal{W}} p(w) \sum_{\substack{q\in\mathcal{Q}, \mathbf{y}\in\mathcal{Y}^{(n)}: \\ h(\mathbf{y},q)\neq w, \mathbf{x}=g(w)}} p(q|w) p(\mathbf{y}|\mathbf{x}) d(w, h(\mathbf{y}, q)), \tag{54}$$

leading to the encoding rule $g(w) = \mathbf{x}$ if

$$f_e(w, \mathbf{x}) \leq f_e(w, \mathbf{x}') \tag{55}$$

for all $\mathbf{x}' \in \mathcal{X}^{(n)}$, where

$$f_e(w, \mathbf{x}') = \sum_{\substack{q\in\mathcal{Q}, \mathbf{y}\in\mathcal{Y}^{(n)}: \\ h(\mathbf{y},q)\neq w, \mathbf{x}'=g(w)}} p(q|w) p(\mathbf{y}|\mathbf{x}') d(w, h(\mathbf{y}, q)). \tag{56}$$

The optimal encoding and decoding functions are the set of $(g,h) \in \mathcal{G} \times \mathcal{H}$ pairs that minimize $D(g,h,p)$ from all the pairs that simultaneously satisfy (52) and (55).

V. Mixed Strategy Nash Equilibrium

For the semantic communication game, a mixed strategy Bayesian Nash equilibrium always exists, which follows from the fact that both player 1 and player 2 have a finite set of possible types and strategies [9], [10]. We next characterize the structure of the mixed strategies at equilibrium.

Let $s_\theta \in \Delta(\mathcal{P})$ be a mixed strategy for player 1 if it is of type θ, where $\Delta(\mathcal{P})$ is the set of all probability distributions over $\mathcal{P}$. In essence, s_θ assigns a probability to each $p \in \mathcal{P}$, which we denote by $s_\theta(p)$, such that $s_\theta(p) \geq 0$ for all $p \in \mathcal{P}$, and $\sum_{p\in\mathcal{P}} s_\theta(p) = 1$. Similarly, let $s \sim \Delta(\mathcal{G} \times \mathcal{H})$ be a mixed strategy for player 2, where $\Delta(\mathcal{G} \times \mathcal{H})$ is the set of all probability distributions over $\mathcal{G} \times \mathcal{H}$. Here s assigns a probability to each $(g,h) \in \mathcal{G} \times \mathcal{H}$, which we denote by $s(g,h)$, such that $s(g,h) \geq 0$ for all $(g,h) \in \mathcal{G} \times \mathcal{H}$, and $\sum_{(g,h)\in\mathcal{G}\times\mathcal{H}} s(g,h) = 1$. The expected payoff for player 1 is

$$\mathbb{E}[u_1(s_\theta, s, \theta)] = \sum_{p_\theta\in\mathcal{P}} \sum_{(g,h)\in\mathcal{G}\times\mathcal{H}} s_\theta(p_\theta) s(g,h) u_1(p_\theta, (g,h), \theta)$$

for type $\theta \in \Theta$, where $u_1(p_\theta, (g,h), \theta)$ is from (10). If player 1 has type $\theta = a$, the expected payoff becomes $\mathbb{E}[u_1(s_a, s, a)]$, whereas if it has type $\theta = h$, the expected payoff becomes $\mathbb{E}[u_1(s_h, s, h)]$. The expected payoff for player 2 is

$$\mathbb{E}[u_2(s, s_\theta, \theta)] = \alpha \sum_{p_a\in\mathcal{P}} \sum_{(g,h)\in\mathcal{G}\times\mathcal{H}} s_a(p_a) s(g,h) u_2((g,h), p_a, a) + (1-\alpha) \sum_{p_h\in\mathcal{P}} \sum_{(g,h)\in\mathcal{G}\times\mathcal{H}} s_h(p_h) s(g,h) u_2((g,h), p_h, h). \tag{57}$$

Then, (s_a^*, s_h^*, s^*) is a mixed strategy Bayesian Nash equilib-

TABLE I
SEMANTIC ERROR COMPARISONS OF ALGORITHM 1 WITH EXHAUSTIVE SEARCH FOR $|\mathcal{W}| = 4$.

$D(g,h,p)$	$\rho = 0.001$	$\rho = 0.01$	$\rho = 0.1$
Simulated Annealing	2.4649×10^{-6}	2.4479×10^{-4}	0.0230
Exhaustive Search	2.4626×10^{-6}	2.4479×10^{-4}	0.0230

rium if and only if

$$s_a^* = \arg\max_{s_a \in \Delta(\mathcal{P})} \sum_{p_a \in \mathcal{P}} \sum_{(g,h) \in \mathcal{G}\times\mathcal{H}} s_a(p_a) s^*(g,h) D(g,h,p_a), \quad (58)$$

$$s_h^* = \arg\min_{s_h \in \Delta(\mathcal{P})} \sum_{p_h \in \mathcal{P}} \sum_{(g,h) \in \mathcal{G}\times\mathcal{H}} s_h(p_h) s^*(g,h) D(g,h,p_h), \quad (59)$$

$$s^* = \arg\min_{s \in \Delta(\mathcal{G}\times\mathcal{H})} \sum_{(g,h) \in \mathcal{G}\times\mathcal{H}} s(g,h) \Big(\alpha \sum_{p_a \in \mathcal{P}} s_a^*(p_a) D(g,h,p_a) + (1-\alpha) \sum_{p_h \in \mathcal{P}} s_h^*(p_h) D(g,h,p_h) \Big). \quad (60)$$

VI. NUMERICAL RESULTS

In our evaluations, we consider a binary symmetric channel (BSC) with a crossover probability of ρ, i.e., abstraction of binary communication over an additive white Gaussian (AWGN) channel with bit error probability ρ, between the encoder and the decoder. We focus on fixed length binary vectors of length n for the channel input and outputs, leading to the channel transition probability

$$p(\mathbf{y}|\mathbf{x}) = \rho^{l(\mathbf{y},\mathbf{x})}(1-\rho)^{n-l(\mathbf{y},\mathbf{x})}, \quad (61)$$

where $l(\mathbf{y},\mathbf{x})$ is the Hamming distance between $\mathbf{y}$ and $\mathbf{x}$. We consider a set of two contexts $\mathcal{Q} = \{$things that originate from a non-living being, things that originate from a living being$\}$, for which we use the shorthand notation *non-living* and *living*, respectively. We select the words for our evaluations from a benchmark set of words utilized in the semantic similarity literature [11].

To calculate the semantic similarities, we use the WordNet interface of the NLTK language processing tool [12]. We use the path-based semantic similarity measure which returns similarity values in the range 0 to 1 that is inversely proportional to the length of the shortest path between two word meanings in the WordNet taxonomy [6]. The similarity between two words is defined as the maximum of all the similarity values between the different meanings the two words can take. This is a corpus-independent similarity measure which allows us to avoid biasing our results in favor of a specific corpus.

In order to gain insight on the structure of the semantic error minimizing encoding and decoding functions, we initially fix the agent's strategy by letting $\mathcal{P} = \{p\}$. We note that finding the semantic-error minimizing encoder and decoder is computationally intensive, following the index assignment arguments in [8]. Therefore, for the algorithm of player 2, we propose the probabilistic metaheuristic in Algorithm 1 based on simulated annealing [13]. Its main idea is to, starting from an initial state, i.e., encoder/decoder assignment, *perturb* the state at each round by modifying the assignment. The new assignment is kept if it performs better than the old one, otherwise, is kept with a probability depending on a temperature parameter. The temperature is reduced gradually, making the fluctuations

Algorithm 1 Simulated Annealing to Minimize Semantic Error

Choose an initial state by assigning g and h to an arbitrary element of $\mathcal{G}$ and $\mathcal{H}$, and calculate $D(g,h,p)$ from (8).
Initialize the melting temperature T_m and the freezing temperature T_f.
Initialize the maximum number of iterations N_{max}.
$T := T_m$.
$N := 1$.
while $T > T_f$ **or** $N < N_{max}$
  $N := N + 1$
  $g_{temp} := g$; $h_{temp} := h$.
  Generate a Bernoulli random variable $K \sim Bern(1/2)$.
  if $K = 1$ **then** Choose a $w \in \mathcal{W}$ uniformly at random and assign $g_{temp}(w)$ to a new random codeword from $\mathcal{X}^{(n)}$.
  else Choose $q \in \mathcal{Q}$ and $\mathbf{y} \in \mathcal{Y}^{(n)}$ uniformly at random and assign $h_{temp}(\mathbf{y}, q)$ to a new random word from $\mathcal{W}$. ▷ Perturb state.
  Calculate $D(g_{temp}, h_{temp}, p)$ from (8).
  $\Delta D := D(g_{temp}, h_{temp}, p) - D(g,h,p)$
  if $\Delta D < 0$ **then** $g := g_{temp}$; $h := h_{temp}$
  else
    ($g := g_{temp}$; $h := h_{temp}$) with probability $e^{-\Delta/T}$
    Reduce temperature.
return g and h

TABLE II
AVERAGE SEMANTIC ERROR FROM ALGORITHM 1 WITH $|\mathcal{W}| = 20$.

	$\rho = 0.001$	$\rho = 0.01$	$\rho = 0.1$
$D(g,h,p)$	0.002501	0.024784	0.224505

less random as the algorithm progresses. The algorithm stops when the temperature reaches a minimum. To investigate how Algorithm 1 performs in practice, we compare it with an exhaustive search method to find the globally optimal encoder and decoder assignments for the following set of words,

$$\mathcal{W} = \{\text{car}, \text{automobile}, \text{bird}, \text{crane}\}, \quad (62)$$

where we set $n = 3$ and $\mathcal{X}^{(n)} = \mathcal{Y}^{(n)} = \{0,1\}^3$. We let $p(\text{non-living}|\text{car}) = p(\text{non-living}|\text{automobile}) = 1$, $p(\text{living}|\text{car}) = 1$, and $p(\text{non-living}|\text{crane}) = 0.5$, noting that: i) $p(\text{non-living}|w) = 1 - p(\text{living}|w)$ for all $w \in \mathcal{W}$, ii) the word *crane* is meaningful in both contexts, where it takes the meaning of an *object that lifts and moves heavy objects* in the context non-living, and a *large bird with a signature long neck* in the context living [7]. The same does not apply for the remaining words, however. For instance, *car* and *automobile* are irrelevant in the context living, whereas *bird* is irrelevant in the context non-living.

We provide the semantic error values evaluated from Algorithm 1 with $T_m = 10$, $T_f = 2.5 \times 10^{-8}$, and $N_{max} = 50000$ in Table I, compared with the exhaustive search results. We observe from Table I that the performance of Algorithm 1 is close to the global optimum obtained by exhaustive search.

Using Algorithm 1, we next determine the encoding and decoding policies for a larger set of words

$$\mathcal{W} = \{\text{car}, \text{automobile}, \text{gem}, \text{jewel}, \text{coast}, \text{shore}, \text{stove}, \text{food}, \text{fruit}, \text{bird}, \text{forest}, \text{monk}, \text{brother}, \text{magician}, \text{crane}, \text{journey}, \text{voyage}, \text{furnace}, \text{noon}, \text{midday}\}, \quad (63)$$

where we let $p(\text{non-living}|w) = 1$ for $w \in \{$gem, jewel, coast, stove, journey, voyage, furnace, noon, midday$\}$ and $p(\text{living}|w) = 1$ for $w \in \{$food, fruit, forest, monk, brother, magician$\}$. We select $n = 4$ and $\mathcal{X}^{(n)} = \mathcal{Y}^{(n)} = \{0,1\}^4$. The semantic error values are given in Table II. Table III demonstrates the corresponding codeword assignments. Ob-

TABLE III
ENCODING FUNCTION $g(w)$ FROM ALGORITHM 1 FOR $w \in \mathcal{W}$.

$g(w)$	$\rho = 0.001$	$\rho = 0.01$	$\rho = 0.1$
$w =$ car	1100	0011	0000
$w =$ automobile	1100	0011	0000
$w =$ gem	0001	0101	1011
$w =$ jewel	0001	0101	1011
$w =$ coast	1010	1010	1110
$w =$ shore	0111	0000	1111
$w =$ stove	0100	1000	1100
$w =$ food	1010	1111	0111
$w =$ fruit	0011	1001	0001
$w =$ bird	1011	1010	0011
$w =$ forest	0010	1101	0000
$w =$ monk	0101	0010	1000
$w =$ brother	1100	0000	1100
$w =$ magician	1111	0100	1101
$w =$ crane	1000	1100	0100
$w =$ journey	0011	0001	1001
$w =$ voyage	0010	0110	1101
$w =$ furnace	0110	1001	1010
$w =$ noon	1111	1111	0111
$w =$ midday	1011	1111	0111

serve that semantically close words, e.g., *car* and *automobile*, are assigned to the same codewords, as well as *gem* and *jewel*. Several semantically distant words are also assigned to the same codewords, such as *car* and *brother* when $\rho = 0.001$ or *magician* and *voyage* when $\rho = 0.1$. The reason behind this is that these words never occur in the same context, hence the decoder can use its context information to distinguish them.

In the second part of our numerical evaluations, we implement a Bayesian game by letting $\alpha = 0.2$, $\rho = 0.1$, and

$$\mathcal{W} = \{\text{car, automobile, coast, shore, bird, monk, crane}\}, \quad (64)$$

where we let $n = 3$, $\mathcal{X}^{(n)} = \{000, 001, 110, 101, 111\}$ and $\mathcal{Y}^{(n)} = \{0, 1\}^3$. To reduce the dimensionality of the strategy spaces, for each given (pure) encoding strategy, we fix the decoder to the minimum error decoder described in (52). In that sense, the coder now has to decide on the probability distribution only over the encoding functions, instead of both the encoder and the decoder, i.e., $s^*(g, h) = s^*(g)$. We define $\mathcal{P} = \{p_1, p_2\}$ such that $p_1(\text{non-living}|w) = p_2(\text{non-living}|w) = 1$ for $w \in \{\text{car, automobile, coast, shore}\}$ and $p_1(\text{living}|w) = p_2(\text{living}|w) = 1$ for $w \in \{\text{bird, monk}\}$. Lastly, we let $p_1(\text{non-living}|\text{crane}) = 0.2$ but $p_2(\text{non-living}|\text{crane}) = 0.8$. We evaluate the mixed strategy Bayesian Nash equilibrium by using the game theory tool Gambit [14]. In our results, we have observed six Nash equilibrium points, each one suggesting a pure strategy Nash equilibrium. In all equilibrium points, the agent's strategy s^*_θ for $\theta \in \{a, h\}$ is such that $s^*_\theta(p_1) = 1$ if $\theta = a$ and $s^*_\theta(p_2) = 1$ if $\theta = h$. The coder's strategy s^* is such that $s^*(g) = 1$ for the encoding functions $g = g(w)$ given in Table IV for the six equilibrium points, respectively, whereas $s^*(g) = 0$ for all the remaining functions in $\mathcal{G}$.

We observe that in all equilibrium points, the synonyms *car* and *automobile* are assigned to the same codewords. Similarly, semantically close words *coast* and *shore* are assigned to close codewords, with a Hamming distance of 1 between them. On the other hand, the Hamming distance between *shore* and *car* is 2, the Hamming distance between *coast* and *car* is 3. For the words in context *living*, we have three possible words,

TABLE IV
NASH EQUILIBRIUM (NE) STRATEGIES FOR THE CODER.

g(w)	car	automobile	coast	shore	bird	monk	crane
NE 1	000	000	111	110	001	110	101
NE 2	000	000	111	110	111	000	101
NE 3	000	000	111	101	111	000	110
NE 4	110	110	001	000	001	110	101
NE 5	110	110	001	000	111	000	101
NE 6	110	110	001	101	001	110	000

bird, *crane*, and *monk*. We observe that the distance between the codewords assigned to *monk* and *crane* is 2 and between *monk* and *bird* it is 3. On the other hand, for *bird* and *crane*, which are semantically closer words in meaning, the Hamming distance between their assigned codewords is 1. Hence, for the words that are meaningful in the same context, whether *living* or *non-living*, semantically close words are assigned to close codewords, with a small Hamming distance between them.

VII. Conclusion

We have considered the transmission of a source that carries a meaning through a noisy channel. An external entity can influence the decoder, whose true characteristic, e.g., friend or foe, is unknown to the communicating parties. We have formulated the semantic communication problem as a Bayesian game and characterized the pure and mixed strategy Bayesian Nash equilibria. Our results indicate that taking into account the semantic distance of transmitted words improves communication performance of intended meanings even in the presence of influencing agents with potentially adversarial actions. Future directions include communication of phrases and complex network structures.

References

[1] G. Bacci, S. Lasaulce, W. Saad, and L. Sanguinetti, "Game theory for networks: A tutorial on game-theoretic tools for emerging signal processing applications," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 94–119, 2016.

[2] Y. Liu, C. Comaniciu, and H. Man, "A Bayesian game approach for intrusion detection in wireless ad hoc networks," in *ACM Workshop on Game Theory for Comm. and Networks*, 2006, p. 4.

[3] Y. Noam, A. Leshem, and H. Messer, "Competitive spectrum management with incomplete information," *IEEE Trans. on Signal Processing*, vol. 58, no. 12, pp. 6251–6265, 2010.

[4] G. He, M. Debbah, and E. Altman, "A Bayesian game-theoretic approach for distributed resource allocation in fading multiple access channels," *EURASIP J. Wireless Comm. and Networking*, vol. 2010, p. 8, 2010.

[5] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," in *Int. Joint Conf. on Artificial Intelligence*, vol. 1, Montreal, CA, Aug. 1995, pp. 448–453.

[6] R. H. Mili, E. Bicknell, and M. Bletner, "Development and application of a metric on semantic nets," *IEEE Trans. on Systems, Man. and Cybernetics,*, vol. 19-33, no. 1, pp. 17–30, Nov. 1989.

[7] G. A. Miller, "WordNet: A lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[8] B. Guler and A. Yener, "Semantic index assignment," in *IEEE Int. Conf. on Pervasive Comp. and Comm. Workshops*, 2014, pp. 431–436.

[9] M. J. Osborne and A. Rubinstein, *A course in game theory*, 1994.

[10] A. Ozdaglar, "Game theory with engineering applications lecture 17: Games with incomplete information: Bayesian Nash equilibria," 2010, http://ocw.mit.edu.

[11] G. A. Miller and W. G. Charles, "Contextual correlates of semantic similarity," *Language and Cog. Processes*, vol. 6, no. 1, pp. 1–28, 1991.

[12] S. Bird, "NLTK: the natural language toolkit," COLING/ACL on Interactive Presentation Sessions, pp. 69–72, 2006, http://www.nltk.org.

[13] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, May 1983.

[14] R. D. McKelvey, A. M. McLennan, and T. L. Turocy, "Gambit: Software tools for game theory," 2014, http://www.gambit-project.org.