



EEG Evidence for Game-Theoretic Model to Ambiguous Pronoun Resolution^{*}

Mengyuan Zhao¹[0000–0002–8917–0780], Zhong Yin^{2,3}[0000–0003–2013–4009], and Ziming Lu³[0000–0002–6215–3745]

¹ School of Marxism, University of Shanghai for Science and Technology, Shanghai 200093, PR China

mengyuan.zhao@usst.edu.cn

² Engineering Research Center of Optical Instrument and System, Ministry of Education, Shanghai Key Lab of Modern Optical System, University of Shanghai for Science and Technology, Shanghai 200093, PR China

yinzhong@usst.edu.cn

³ Department of Linguistics and Translation, School of International Studies, Zhejiang University, Hangzhou, 310058, PR China

ziminglu@zju.edu.cn

Abstract. In this paper we develop a game-theoretic model of ambiguous pronoun resolution, namely, the pronoun reference is not clearly determined in the context. We propose that iterated best response (IBR) reasoning offers a reasonable solution to ambiguous pronoun processing. Using electroencephalogram (EEG) (14 channels) to investigate Chinese processing, we provide evidence that the processes of resolving ambiguous and unambiguous pronouns are significantly different at both neural and behavioural level. The differences mainly manifest in longer reaction time and signals collected from the channels O1 (left occipital cortex) and P8 (right inferior parietal cortex), which are activated during probabilistic expected utility generation. These findings are consistent with general assumptions of our model that ambiguous pronoun resolution involves a mechanism of decision-making.

Keywords: Pronouns · Game theory · EEG.

1 Introduction

Personal pronouns such as *he* and *she* refer to an earlier mentioned person in the context. Pronoun resolution is a fundamental process in daily language processing, and many linguistic studies have investigated how people assign a referent to a pronoun according to grammatical rules (see [1–4]). These researches have shown that some types of linguistic cues can be used to pronoun resolution including: gender, verb-bias, focus

^{*} Supported by the National Social Science Fund under Grant No. 18CZX014, the Shanghai Philosophy and Social Sciences Fund under Grant No. 2017EZX008, the National Natural Science Foundation of China under Grant No. 61703277, Shanghai Sailing Program under Grant No. 17YF1427000. This chapter is in its final form and it is not submitted to publication anywhere else.

and so on. However, pronouns carry little content by nature, and thus are referentially ambiguous in certain cases. Compare, for example, the following sentences:

- (1) The wife stopped the husband. She cried.
- (2) The owner blamed the waiter. He was angry.

In (1), the gender information of the nouns help to identify pronoun denotation, and therefore the pronoun *she* must refer to *the wife*. In this case, the pronoun's referent is determined by some linguistic cues (here, gender), and we call it *the unambiguous pronoun*. While in (2), gender is not informative to determine the pronoun's referent, we call it *the ambiguous pronoun*. The pronoun *he* in (2) may either refer to *the owner* or *the waiter*, and its referent is undetermined. Another linguistic cue that may help pronoun resolution is the verb-bias, that is, the semantic meaning of the verb may lead to a bias for pronoun interpretation. In a seminal work of psycholinguistics on pronouns, Garvey & Caramazza [1] have reported strong bias in pronoun resolution for specific verbs. They found a strong bias in interpreting *she* as a reference of the object *the daughter* for the verb *scold* (e.g. *The mother scolded her daughter because she ...*), while *she* referring to the subject *the mother* for the verb *confess* (e.g. *The mother confessed to her daughter because she ...*). This suggests that people do not simply decide the referent of an ambiguous pronoun by proximity.

Unambiguous expressions show *prima facie* advantages comparing to their ambiguous counterparts. Actually, the pronoun *he* as shown in sentence (2) is more brief than the definite descriptions *the owner* or *the waiter*. Brevity is commonly argued as a rationality principle in communication. Grice's [5] Maxims of Conversation have shown this tension between brevity and ambiguity: Maxim of Quantity and Maxim of Manner. Therefore, it is worthy of discussion about the rationale of the use of ambiguous expressions in daily communication. The present work assumes that ambiguous pronoun processing involves rational decision-making, which can be modelled in game theory. The tradition of applying game models to pragmatics can be traced back to the seminal work of David Lewis [6]. Lewis introduced signalling games, which characterize communication as a speaker's attempt to influence a hearer's action by sending a certain signal. On the basis of Lewisian tradition, Parikh [7, 8] developed a more comprehensive framework named *games of partial information*. Parikhian model has been applied to analyze reference resolution (see [9, 10]). Clark and Parikh [10] adopted the conception of *Pareto Nash Equilibrium* to solve the game of ambiguous reference. Another branch of game-theoretic pragmatics couches iterative dynamics as an analysis of rational language use (see for example [11–13]). An influential model of iterated response reasoning is the so called *pragmatic back-and-forth reasoning* developed by Franke and Jäger [13]. Recently, the dynamic reasoning model has been applied to an analysis of ambiguous expressions [14].

In this paper, we develop a game-theoretic model of ambiguous pronoun resolution. The model includes two parts: a signalling game as an illustration of the situations where people process ambiguous pronouns, and a reasoning account as a solution of the game. We point out that the solution conception of Parikhian model requiring the agents being rational enough to select the most profitable strategies is too constrictive. To solve the game of ambiguous pronoun, we introduce the iterated best response (IBR)

reasoning. We argue that IBR reasoning by assuming a step-by-step interactive reasoning procedure allows analyzing actions under bounded rationality.

It is arguable whether the cognitive processing of pronoun resolution only involves the core language network or it also recruits the network of strategic decision-making as suggested in our model. To test the assumptions of our model, we use electroencephalogram (EEG) (14 channels) to study ambiguous pronoun processing in Chinese. Though recent years some work has been done on neural measures of reference resolution (see [15–19]), neural correlates to ambiguous pronoun resolution remain greatly uncharacterized. In our experiments, participants observed a sentence containing two nouns followed by the other sentence containing a personal pronoun (for example, sentence (1)). Behavioural data including reaction time (the time consumed by a participant to identify a pronoun referent) and neural data including EEG signals of 4 frequency bands and 14 channels during the whole resolution procedures are recorded. These data suggest a significant difference between processes of ambiguous and unambiguous pronoun resolution. The neural data of channel P8 and O1 demonstrate more activated recruitment of right inferior parietal cortex and left occipital cortex during ambiguous pronoun resolution. According to a review of neuroimaging studies (see [20, 21, 18]), these areas respectively implicate expected utility calculation under probabilistic situations and extra effort paid for entire sentence reading. These findings are consistent with the assumptions of our model that ambiguous pronoun resolution involves a mechanism of decision-making. Our work also extends previous research on English processing to Chinese processing, which is structurally different from English. The results indicate that game-theoretic model can be applied across languages, and encourage further generalization of the model in future research.

2 A Game-Theoretic Model to Ambiguous Pronoun Resolution

In this section, we will first construct a game-theoretic model, and then apply it to the case of ambiguous pronoun resolution. The model follows the main assumptions from the tradition of Gricean pragmatics (see [5]): communication is considered to be a cooperative and rational activity. The model presented in this paper will consider about a basic case that involves just one ambiguous message. The current model is aiming at offering a brief guideline to analyze ambiguous pronoun resolution and is competent to explain the rationale of the pragmatics in cases such as sentence (2). To analyze more complicated sentences with more than one ambiguous pronouns would require an extended model which can be derived from the basic model in principle. However, the development of an extended model has gone beyond the present work.

2.1 The Model

Context Modelling In the model, we assume that there is a speaker S , who has the relevant information about the world where she is in, and a hearer H , who has to judge about the world by reasoning on the message that the speaker transmits. Assuming there are two possible worlds: w_1 and w_2 , we now model the speaker’s knowledge about the world as types: t_1 indicates that S knows that she is in w_1 , t_2 indicates that S knows

that she is in w_2 . We introduce Nature, say N , as an impersonal player of the game. N chooses a move to either type, say $t \in T$, with a probability, say p . Let $Pr(t) \in \Delta(T)$ be prior distribution over types, where $\Delta(T)$ refers to a probability distribution over types t_1, t_2, \dots, t_n . We assume that $p = Pr(t)$ implicates H 's prior belief in t before receiving any message. S will send a message, say $m \in M$, to H to inform her about the world, and H will interpret the received message into a type of S . We assume that the players will play according to the semantic meaning of messages. This constraint can be shown by introducing a lexicon B that maps type-message pair to the Boolean truth-value of the message for the speaker's type. A minimal lexicon fragment that involves choices between ambiguous and unambiguous messages is one with two types and three messages, where $B(t_1, m_1) = B(t_1, \bar{m}) = 1$, $B(t_2, m_2) = B(t_2, \bar{m}) = 1$. Put into words, for type t_1 , S may either send an unambiguous message containing a definite description, say m_1 , or send an ambiguous message containing a pronoun, say \bar{m} . Similarly, for type t_2 , S may send an unambiguous message m_2 or an ambiguous message \bar{m} . Accordingly, we assume that the speaker will play a semantically consistent strategy, say σ , which is defined as follows:

Definition 1. A semantically consistent strategy of the speaker σ is a function that maps each speaker type $t \in T$ to a probability of each message $m \in M$ to be sent in t , given that m is semantically true in t : $\sigma \in \Delta(M)^T$, $B(t, m) = 1$.

Thereafter, H will interpret m_1 into t_1 , m_2 into t_2 . And when H receives \bar{m} , she may possibly interpret it into either m_1 or m_2 . Accordingly, we assume that the hearer will also play a semantically consistent strategy, say ρ , which is defined as follows:

Definition 2. A semantically consistent strategy of the hearer ρ is a function that maps each message m to a probability of each interpretation t to be chosen, given that m is semantically true in t : $\rho \in \Delta(T)^M$, $B(t, m) = 1$.

We further assume that a successful communication using a pronoun will provide players with an extra gain, say ε , since the communication is complete with less words. We assume that S and H are purely cooperative, in the sense that S and H share the same utility functions. It means that both S and H would prefer that H 's interpretation of m , say t_j , is identical to S 's type, say t_i . We define the utility functions of players as follows.

Definition 3. Let $U_N(t_i, m, t_j)$ be payoff of $N \in \{S, H\}$ given t_i , m and t_j , where $i, j = 1, 2$.

$$U_N(t_i, m, t_j) = \begin{cases} 1, & \text{if } i = j \text{ and } m \in \{m_1, m_2\} \\ 1 + \varepsilon, & \text{if } i = j \text{ and } m = \bar{m} \\ 0, & \text{if } i \neq j \end{cases}.$$

Definition 3 suggests: Both players will gain a plain payoff, say 1, using unambiguous messages; both will gain an extra payoff, say $1 + \varepsilon$, if H correctly understands the ambiguous pronoun message; and both will earn 0, if H misinterprets the pronoun. We denote by $p \in (0, 1)$ H 's prior belief that S is of type t_1 , i.e. $Pr(t_1) = p$. The game tree in Fig. 1 illustrates the signalling game of our model.

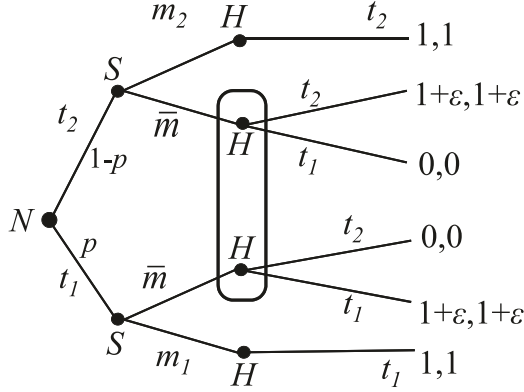


Fig. 1. A game tree for the model.

Solution Modelling As a solution of the game, we introduce the IBR reasoning framework. The IBR reasoning includes two reasoning sequences, namely, the S_0 -sequence and the H_0 -sequence. In the S_0 -sequence, the first step of reasoning starts from a level-0 speaker, say S_0 . We assume that S_0 is a naïve speaker, in the sense that she will choose a random message that is semantically consistent with her type. The level-1 hearer, say H_1 will play rationally based on her belief in S_0 , in the sense that H_1 will choose the strategy that offers her the best expected utility. Similarly, S_2 will play rationally based on her belief in H_1 , and the reasoning continues in this way. In the H_0 -sequence, the first step will be taken by a level-0 hearer, say H_0 . H_0 will choose the strategy that offers her the best expected utility based on her posterior belief in the speaker's strategy. In other words, we assume that H_0 , unlike S_0 , is rational. The level-1 speaker, say S_1 will play rationally according to her belief in H_0 , and so on. As a generalization, a level- $(k+1)$ player will play rationally according to her belief in the strategies that a level- k player will choose. Now we illustrate by induction the IBR reasoning scaffolding.

Naïve Levels The S_0 -sequence starts from the level-0 speaker S_0 , who will randomly play a semantically consistent strategy. According to Definition 1, S_0 may choose either m_1 or \bar{m} in t_1 , and she may choose either m_2 or \bar{m} in t_2 . We perspicuously list all possible choices of S_0 for both types as follows.

$$S_0 = \left\{ \begin{array}{l} t_1 \mapsto m_1, \bar{m} \\ t_2 \mapsto m_2, \bar{m} \end{array} \right\}.$$

The H_0 -sequence starts from the level-0 hearer H_0 , who will play rationally according to her posterior belief. A posterior belief of H_0 , say μ_0 , results from the hearer's prior belief updated by the semantic meaning of the messages: $\mu_0(t|m) = Pr(t) \times B(t, m)$. For unambiguous messages m_1 and m_2 , H_0 will choose semantic interpretations t_1 and t_2 , respectively. For the ambiguous message \bar{m} , the posterior beliefs in two types are calculated as follows: $\mu_0(t_1|\bar{m}) = Pr(t_1) \times B(t_1, \bar{m}) = p$, $\mu_0(t_2|\bar{m}) = Pr(t_2) \times B(t_2, \bar{m}) = 1 - p$. H_0 will choose any t with a higher posterior belief as an interpretation of \bar{m} . This means H_0 's interpretation of the ambiguous message is dependent on the value of p , which represents the hearer's prior belief in speaker's types.

$$H_0 = \left\{ \begin{array}{l} \left\{ \begin{array}{l} m_1 \mapsto t_1, \\ m_2 \mapsto t_2, \end{array} \text{ if } p > \frac{1}{2} \right. \\ \left. \begin{array}{l} \bar{m} \mapsto t_1, \\ m_1 \mapsto t_1, \\ m_2 \mapsto t_2, \\ \bar{m} \mapsto t_2, \end{array} \text{ if } p < \frac{1}{2} \right. \end{array} \right\}.$$

Sophisticated Levels We assume that in both S_0 - and H_0 -sequences, level- $(k+1)$ player will play as the best response to her belief in the other's strategies ($k > 0$). For simplicity, we assume that the players believe that their opponents are of the level that is exactly one level lower than her own. For the level- $(k+1)$ speaker, her belief is given as the semantically consistent strategies (see Definition 2) of the level- k hearer, say ρ_k . We further assume that the speaker will give a best response to her belief, say $Br(\rho)$. A best response means a rational play, namely, the speaker will choose a pure strategy, say $s(t)$, that will maximize her expected utility. The speaker's expected utility, say $EU_S(t, m, \rho)$, can be calculated as follows.

$$EU_S(t, m, \rho) = \sum_{t_i \in T} \rho(t_i|m) \times U_S(t, m, t_i) \quad (1)$$

As shown in Equation (1), the expected utility of the speaker is a sum of the utilities (see Definition 3) of all possible outcomes weighted by their probabilities dependent on the speaker's belief. Accordingly, the speaker's strategy as a best response of her belief is as follows.

$$s(t) = BR(\rho) \in \arg \max_{m \in M} EU_S(t, m, \rho) \quad (2)$$

As for the level- $(k+1)$ hearer, her belief is given as the posterior belief, say μ_{k+1} , which is derived from the hearer's prior belief and the semantically consistent strategies (see Definition 1) of the level- k speaker, say σ_k , by Bayesian conditionalization.

$$\mu_{k+1}(t_j|m_i) = \frac{Pr(t_j) \times \sigma_k(m_i|t_j)}{\sum_{t' \in T} Pr(t') \times \sigma_k(m_i|t')} \quad (3)$$

In Equation (3), the Bayesian conditionalization represents the belief dynamics: the likelihood for each type t is computed after a certain message m is received in t given

prior probability of t and the expected probability of m to be sent in t . We assume that the hearer will show a best response to her belief, say $Br(\mu)$. A best response of a hearer is a pure strategy, say $h(t)$, that will maximize her expected utility. The hearer's expected utility, say $EU_H(t, m, \mu)$, can be calculated as follows.

$$EU_H(t, m, \mu) = \sum_{t_i \in T} \mu_1(t_i|m) \times U_H(t_i, m, t) \quad (4)$$

In Equation (4), the hearer's expected utility is a sum of the utilities (see Definition 3) of all possible outcomes weighted by their probabilities dependent on the hearer's belief. Accordingly, the hearer's strategy as a best response of her belief is as follows.

$$h(m) = BR(\mu) \in \arg \max_{t \in T} EU_H(t, m, \mu) \quad (5)$$

Now we start reasoning in the S_0 -sequence. After S_0 sends a certain m to H_1 , the latter will respond based on her posterior belief in S_0 , say $\mu_1(t|m)$. From Equations (3) and (4), we calculate H_1 's expected utilities of different choices while receiving \bar{m} : $EU_{H_1}(t_1, \bar{m}, \mu_1) = p \times (1 + \varepsilon)$, $EU_{H_1}(t_2, \bar{m}, \mu_1) = (1 - p) \times (1 + \varepsilon)$. From equation (5), H_1 will interpret \bar{m} into t_1 if and only if $EU_{H_1}(t_1, \bar{m}, \mu_1) > EU_{H_1}(t_2, \bar{m}, \mu_1)$, requiring $p > \frac{1}{2}$. Thus, the strategy of H_1 can be perspicuously illustrated as follows.

$$H_1 = \left\{ \begin{array}{l} \left\{ \begin{array}{l} m_1 \mapsto t_1, \\ m_2 \mapsto t_2, \end{array} \text{ if } p > \frac{1}{2} \right. \\ \left. \left\{ \begin{array}{l} \bar{m} \mapsto t_1, \\ m_1 \mapsto t_1, \\ m_2 \mapsto t_2, \\ \bar{m} \mapsto t_2, \end{array} \text{ if } p < \frac{1}{2} \right. \right. \end{array} \right\}.$$

Accordingly, $h_1(m)$ can be calculated from Equation (5): $h_1(t_1|m_1) = h_1(t_2|m_2) = h_1(t_1|\bar{m}) = 1$, if $p > \frac{1}{2}$; $h_1(t_1|m_1) = h_1(t_2|m_2) = h_1(t_2|\bar{m}) = 1$, if $p < \frac{1}{2}$. S_2 will respond upon her belief, say ρ_1 , which is equal to h_1 . From equations (4) and (5), we illustrate S_2 's strategies as follows.

$$S_2 = \left\{ \begin{array}{l} \left\{ \begin{array}{l} t_1 \mapsto \bar{m}, \\ t_2 \mapsto m_2, \end{array} \text{ if } p > \frac{1}{2} \right. \\ \left. \left\{ \begin{array}{l} t_1 \mapsto m_1, \\ t_2 \mapsto \bar{m}, \end{array} \text{ if } p < \frac{1}{2} \right. \right. \end{array} \right\}.$$

Accordingly, $s_2(t)$ can be computed from Equation (2): $h_1(\bar{m}|t_1) = h_1(m_2|t_2) = 1$, if $p > \frac{1}{2}$; $h_1(m_1|t_1) = h_1(\bar{m}|t_2) = 1$, if $p < \frac{1}{2}$. H_3 will give the best response upon her posterior belief in S_2 , and her strategy can be figured out following a similar procedure of what happens in the case of H_1 . We perspicuously illustrate H_3 's strategy as follows.

$$H_3 = \left\{ \begin{cases} m_1 \mapsto t_1, \\ m_2 \mapsto t_2, & \text{if } p > \frac{1}{2} \\ \bar{m} \mapsto t_1, \\ m_1 \mapsto t_1, \\ m_2 \mapsto t_2, & \text{if } p < \frac{1}{2} \\ \bar{m} \mapsto t_2, \end{cases} \right\}.$$

It is clear that H_3 shows the same strategic pattern as H_1 . Given the reasoning principles of best response operation, S_4 should play the same as S_2 , H_5 should play the same as H_3 and so on. This means the strategies of the players start to repeat themselves from the level-3 hearer after two rounds of best response reasoning in the S_0 -sequence. And it is also easy to reach similar results in the H_0 -sequence, of which we would like to skip the details for the sake of simplicity.

Predictions In our model, both the sets T and M are finite, therefore, there are finitely many pure strategies. This means the IBR reasoning sequences will definitely repeat themselves at a certain level. We define an idealized prediction of IBR reasoning as follows.

Definition 4. *The idealized predictions of IBR reasoning are infinitely repeated strategies S^* and H^* :*

$$S^* = \{s \in S \mid \exists i \forall j > i : s \in S_j\}$$

$$H^* = \{h \in H \mid \exists i \forall j > i : h \in H_j\}$$

From the steps that we have shown in the S_0 -sequence, the strategy repetition begins after two rounds of reasoning. And it is also easy to prove that a reasoning of H_0 -sequence will lead to similar results. Accordingly, we provide a prediction of our model in Proposition 1:

Proposition 1.

$$S^* = \left\{ \begin{cases} t_1 \mapsto \bar{m}, \\ t_2 \mapsto m_2, & \text{if } p > \frac{1}{2} \\ t_1 \mapsto m_1, \\ t_2 \mapsto \bar{m}, & \text{if } p < \frac{1}{2} \end{cases} \right\}. H^* = \left\{ \begin{cases} m_1 \mapsto t_1, \\ m_2 \mapsto t_2, & \text{if } p > \frac{1}{2} \\ \bar{m} \mapsto t_1, \\ m_1 \mapsto t_1, \\ m_2 \mapsto t_2, & \text{if } p < \frac{1}{2} \\ \bar{m} \mapsto t_2, \end{cases} \right\}.$$

Proposition 1 suggests: Whether an ambiguous pronoun is used instead of an unambiguous noun is dependent on the hearer's prior belief in the frequency of the world where the speaker is in. When the hearer believes that it is more likely for the speaker to be in the world w_1 , where S is of type t_1 , the speaker will send a pronoun message for t_1 and a noun message for t_2 , and the hearer will successfully translate the pronoun message into t_1 . And the same reasoning will also stand in the case where the hearer's prior belief is biased towards t_2 .

2.2 Comparison with other Models

Gricean Approaches The game-theoretic model of ambiguous pronoun resolution presented here is highly in the spirit of Gricean pragmatics (see [5]). Grice accounted for pragmatic reasoning as a rational behaviour of agents. Pronoun resolution has been explained by application of the Gricean or neo-Gricean approaches (see for example [22, 23]). Our game model follows Grice's idea by modelling inferences of ambiguous pronoun resolution as rational interaction between the speaker and the hearer.

The differences between our model and the Gricean approaches are mainly in two aspects. Firstly, in the conceptual aspect, the game model presented in this paper does not rely on a formulation of Grice's Maxim of Conversation. The game model is constructed based on a simple assumption of cooperation in the sense that both of the players share a common interest. And this cooperation is formalized in terms of the utility functions (see Definition 3). Furthermore, our model also leaves open the possibility of explaining non-cooperative situations by a revision of utility functions. Secondly, in the epistemic aspect, the model presented here uses an iterated best response reasoning to show epistemic concerns of which the Gricean approaches are lack. The IBR reasoning have three features: semantic meaning focus, step-by-step interactive pattern and tolerance of bounded rationality. These features correspond to actual epistemic situations. IBR reasoning starts from level-0 players, who select according to the semantic meaning of the messages. The semantic meaning acts as a psychological focus of the agents during the reasoning, that is, the agents are psychologically attracted by the semantic meaning, from which they start the pragmatic reasoning. The IBR model also simulates a step-by-step interactive reasoning. This framework allows agents to update their belief in each other's rational strategies, and to upgrade their reasoning level. In addition, IBR reasoning is tolerant to limited rationality, which shows a more real situation. The model offers freedom to stop at any level of sophistication to check the result of reasoning from either bounded or ideal rationality.

Games of Partial Information A game of partial information involves ambiguous information states in the game tree and is to be solved by adoption of Pareto-Nash Equilibrium (see [7, 8]). It has been applied to the analysis of pronoun resolution (see [9, 10]). Games of partial information share the same tradition of game-theoretic pragmatics with the model presented in this paper.

The difference between the model presented here and the games of partial information is mainly in the aspect of solution concepts. The games of partial information adopt the solution concept of Pareto-Nash Equilibrium. A Pareto-dominant Nash Equilibrium is the best-paid strategy profile among those which offer both players the best payoff given the strategy of her opponent. In other words, the Pareto-Nash Equilibrium is the most profitable equilibrium of the game. To follow this, it is required that the agents compute all equilibria and make the comparison as well. This requirement not only is too much for the rationality of the agents, but also presumes an outsider's view of the game to complete the calculation. In comparison, our model adopts the solution of IBR reasoning predictions. The IBR reasoning illustrates a step-by-step interaction. It allows the agents to respond from different levels of sophistication with more tolerance to the rationality of the players. It also shows as a simulation of the real procedure of

the agents updating their belief and responding to it. Therefore, the IBR reasoning is more like an insider's view of the game.

Pragmatic Back-and-Forth Reasoning The back-and-forth reasoning combines the idea of signalling games as the context formulation and iterated response reasoning as the solution schemes (see [12, 13]). It has been applied to analyze the resolution of ambiguous reference (see [14]). The model presented here bears a close resemblance to the back-and-forth reasoning.

The difference between our model and the back-and-forth reasoning is mainly in two points. Firstly, as for the context modelling, the back-and-forth reasoning uses signalling games to describe the context of a sender sending messages to inform a receiver about the state t . Instead, in our model, we use t to represent the type of the speaker, who will send a message to inform the hearer about her knowledge of the world. Comparing to the settings of the back-and-forth reasoning, our model is capable of representing the speaker's expertise, and thus leaves open the possibility that the speaker has only partial information of the world. Secondly, as for the solution modelling, the back-and-forth reasoning includes at least three types of iterated response reasoning schemas: iterated best response, iterated cautious response and iterated quantal response. Our model adopts a solution concept that is most close to the iterated best response in Franke's work (see [12]). In the vanilla IBR model, Franke assumed that the receiver would show unbiased prior beliefs in all states. In comparison, our work introduces a parameter p to represent the hearer's prior belief in different speaker types. This parameter is key to our model in the sense that it determines the final solution to the game. Furthermore, the parameter of prior belief also plays an important role both in our pretest work and in the experiments (see Section 3 for details).

2.3 The Application

Game-theoretic models have been applied to various pragmatic phenomena (see [24–26] for a selective survey). However, most researches are based on an analysis of English sentences. We explore our game-theoretic model to an analysis of Chinese, which is structurally different from English. We first construct 200 pairs of Chinese sentences dividing into different groups, then we investigate the value of prior belief in the referent of ambiguous pronouns based on a 30-participant survey.

We construct 200 pairs of sentences through the following steps: We first identify 80 nouns, 40 of which are gender-biased (for example, *qizi* 'wife') and the other 40 gender-neutral (for example, *laoshi* 'teacher'); we then generate meaningful 200 sentences by pairing nouns with a transitive verb (for example, *piping* 'blame'); we finally generate 200 sentences including a pronoun and an intransitive verb (for example, *xiao* 'smile'). We divide the 200 pairs of sentences into different classes according to a group of characteristics. A main classification is to distinguish between ambiguous and unambiguous pronoun resolution. For example, compare the following pairs of sentences:

- (3) Qizi lanzhu zhangfu. Ta ku-le.
 wife stop husband. She cry ASP.
 'The wife stopped the husband. She cried.'

- (4) Dianzhang piping fuwuyuan. Ta shengqi-le.
 owner blame waiter. He angry ASP.
 ‘The owner blamed the waiter. He was angry.’

The pronoun in (3) unambiguously refers to *the wife*, while the pronoun in (4) may either refer to *the owner* or *the waiter*. Our model can be applied to analyze ambiguous pronoun resolution as shown in example sentences (4). Since the hearer’s prior belief is key to solve the game as shown in Proposition 1, we conduct a pretest survey to investigate how the verb-bias as a linguistic cue influence people’s resolution to ambiguous pronouns. 30 healthy young adults from University of Shanghai for Science and Technology participated in the survey. We replace the nouns of each sentences with *X* and *Y* (for example, *X piping Y. Ta shengqi-le.* ‘X blamed Y. She was angry.’), and ask the participants whether the pronoun refers to *X* or *Y*. On average, the object noun is preferred (74.5%).

We now apply the model to analyze the case shown in example sentences (4). Here are two possible worlds: w_1 , where the owner was angry, and w_2 , where the waiter was angry. Accordingly, the speaker’s types include: the speaker knows that she is in w_1 , say t_1 , and she knows that she is in w_2 , say t_2 . From the survey analysis, the hearer’s prior belief is biased towards t_2 , that is, $p < \frac{1}{2}$. According to Proposition 1: If the speaker is of t_1 , she will utter *The owner was angry*; if the speaker is of t_2 , she will utter *He was angry*. The hearer, after receiving the message *He was angry*, will interpret it into t_2 , namely, assigning the referent *the waiter* to the pronoun *he*.

The prediction of our analysis on ambiguous pronoun resolution is consistent with the results of our EEG experiment, which will be illustrated in the next section.

3 The Experiment

3.1 Methods

10 healthy adults from University of Shanghai for Science and Technology participated in the EEG experiment. All participants are native speaker of Chinese and right-handed. We excluded one participant due to significantly low accuracy rate of unambiguous pronoun resolution results. Therefore, all data analyses are based on 9 healthy adults.

We first construct 200 pairs of meaningful sentences in Chinese as described in Section 2.3. The sentences include both unambiguous pronouns and ambiguous pronouns. We not only construct sentences with the syntactic structure of $S+V+O$, (e.g. sentences (3) and (4)), but also construct pairs of sentences with unique syntactic structures of Chinese, N_1+N_2+V . Consider, for example, the following sentences:

- (5) Laoshi he yanjiuyuan yuehui. Ta xiao-le.
 teacher and researcher date. He smile ASP.
 ‘The teacher dated the researcher. He smiled.’

The sentences are displayed on a computer screen. For each experimental trial, a fixation cross (500ms) shows first, and then three stimulus events follow (each 3000 ms). The first stimulus event is the presentation of a sentence containing two nouns

(e.g., *Qizi lanzhu zhangfu*. ‘The wife stopped the husband.’). The second stimulus event is the presentation of a sentence containing a pronoun (e.g., *Ta ku-le*. ‘She cried.’). The last stimulus event is a question for participant to choose whether the pronoun refers to the noun on he left (e.g. *qizi* ‘the wife’) or the noun on the right (e.g. *zhangfu* ‘the husband’). Participants are given up to 5500ms to respond, and their responses are recorded by tapping a certain key on the keyboard (key *z* as choosing the left noun, and key *m* for the right noun). All 200 pairs of sentences of either unambiguous classes or ambiguous classes are pseudo-randomly distributed. The experimental procedure is illustrated in Fig. 2.

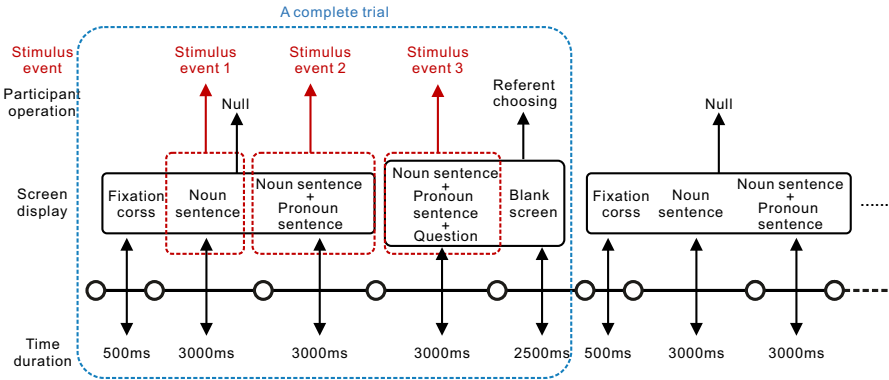


Fig. 2. The experiment procedure

EEG was recorded by Emotiv Xavier SDK at a sampling rate of 128 Hz using 14 Cu electrodes, placing according to international 10-20 system. The frequency bands included in EEG signals are as follows (see [27]): Theta (4-8 Hz), Alpha (8-12.5 Hz), Beta(12.5-28 Hz), Gamma (30-40 Hz). The schematic representation of the 14-channel positions is illustrated in Fig. 3(a).

3.2 Results and Discussion

We evaluate the response accuracy for the unambiguous pronoun resolution. The performance was highly accurate: $M=95.8\%$. These results confirm the efficacy of gender information as useful for pronoun resolution. To evaluate the uses of verb-bias information, we analyze the proportion of participants assigning the referent to preferred object noun in the cases of ambiguous pronoun resolution. As predicted in the pretest survey (see Section 2.3), participants prefer choosing the object as the referent of pronoun ($M=69\%$). We also analyze the cases of ambiguous pronoun resolution with a sentence structure of N_1+N_2+V , and participants show preference to the second noun as the referent of the pronoun ($M=62\%$).

We investigate reaction times of pronoun resolution of various classes. We find that there are significant differences in reaction times between ambiguous and unambigu-

ous pronoun resolution in both sentence structures, namely $S+V+O$ and N_1+N_2+V . More specifically, we report significant differences in reaction times collected from all 9 participants while resolving the following three groups of comparisons: Firstly, comparing unambiguous pronoun identified directly by both gender-biased nouns (e.g., sentences (3)) relative to ambiguous pronoun with the structure of $S+V+O$ (e.g., sentences (4)) shows a p value of t-test much smaller than 0.01 ($p=0.000$); secondly, unambiguous pronoun identified indirectly by a gender-biased noun and a gender-neutral noun (e.g., sentences (6)) versus ambiguous pronoun with the structure of $S+V+O$ (e.g., sentences (4)) ($p=0.000$); thirdly, unambiguous pronoun identified directly by gender-biased nouns (e.g., sentences (8)) versus ambiguous pronoun with the structure of N_1+N_2+V (e.g., sentences (5)) ($p=0.000$).

- (6) Kuaiji guli nüer. Ta xiao-le.
 accountant encourage daughter. he smile ASP.
 ‘The accountant encouraged the daughter. He smiled.’
- (7) Nüyanyuan he zhangfu zhengchao. Ta juezui-le.
 actress and husband quarrel. she pout ASP.
 ‘The actress quarrelled with the husband. She pouted.’

To investigate the neural correlates of pronoun resolution, we compare EEG data collected from 9 participants while processing unambiguous pronouns relative to the data of ambiguous pronoun resolution. More specifically, we report significant difference in EEG data collected from all 9 participants while resolving two groups of comparisons. Firstly, for EEG data collected from channel P8 for frequency bands Theta and Alpha, significant differences ($p=0.005$ for both frequency bands in signrank test) have been reported comparing resolution of unambiguous pronoun identified directly by gender-biased nouns (e.g., sentences (8)) relative to resolution ambiguous pronoun with the structure of N_1+N_2+V (e.g., sentences (5)). This result is illustrated as box plots in Fig. 3(c)-(d), where the significant difference from channel P8 is highlighted in red. Secondly, for EEG data collected from channel O1 for frequency bands Beta, a significant difference ($p=0.006$ in signrank test) has been reported comparing resolution of unambiguous pronoun identified directly by both gender-biased nouns (e.g., sentences (3)) relative to resolution ambiguous pronoun with the structure of $S+V+O$ (e.g., sentences (4)). This result is illustrated as box plots in Fig. 3(e)-(f), where the significant difference from channel O1 is highlighted in red.

The EEG signals from channels P8 and O1 are correlated with the recruitment of right inferior parietal cortex and left occipital cortex (see [28]). Right inferior parietal cortex is a region that has been correlated to an integration of of probabilistic assessment and evaluation of expected utility (see [20, 18]). Activation in occipital cortex has been reported as an increase of brain workload of the entire sentence reading (see [18]), which can be explained as correlated to increased demands of language processing.

4 Conclusions and Outlook

In this paper, we construct a game-theoretic model for ambiguous pronoun resolution. The behavioural findings suggest that people use gender information of nouns for pro-

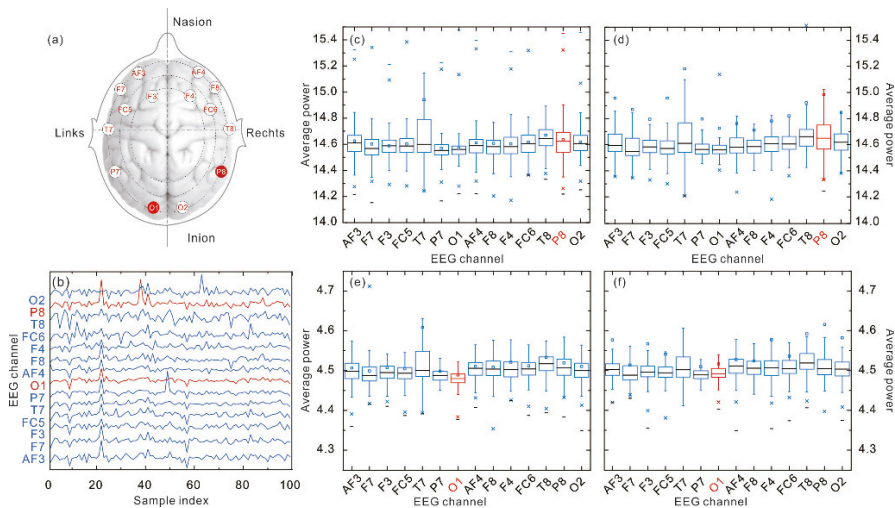


Fig. 3. EEG recording, localization of the electrodes position and signal difference comparing unambiguous to ambiguous pronoun processing. (a) Schematic representation of the 14 electrodes, and signals from channel P8 and O1 highlighted in red show significant difference between unambiguous and ambiguous pronoun processing. (b) EEG signals from 14 channels of a single participant during pronoun processing of 100 pairs of sentences, and signals from channel P8 and O1 are highlighted in red. (c)-(d) Box plots of EEG data of frequency Theta from 14 channels of 9 participants while processing unambiguous and ambiguous pronouns of $S+V+O$ and a significant difference shown in channel P8 is highlighted in red. (e)-(f) Box plots of EEG data of frequency Beta from 14 channels of 9 participants while processing unambiguous and ambiguous pronouns of N_1+N_2+V , and a significant difference shown in channel O1 is highlighted in red.

noun resolution with high accuracy. When gender is not informative, people use verb-bias information to resolve ambiguous pronouns. The experimental results are consistent with the assumptions and predictions of our model in three ways. Firstly, people spending more time in ambiguous pronouns than unambiguous pronouns is consistent with the assumption of our model that ambiguous pronoun resolution involves more complicated and time-consuming cognitive procedures, namely, decision-making. Secondly, people's consistent preference over object nouns in ambiguous pronoun resolution revealed in both pretest surveys and experiments is consistent with the prediction of our model that the solution to the game of ambiguous pronoun resolution is dependent on hearer's prior belief in speaker's types. Thirdly, significant differences shown in EEG channel P8 (right inferior parietal cortex) and O1 (left occipital cortex) comparing ambiguous to unambiguous pronoun processing implicates that the evaluation of probabilistic expected utilities are involved and accordingly increased demands of sentence processing are also involved, which is consistent with the assumption of our model.

The experimental results shown in this paper offers evidence that the epistemic processing of ambiguous pronoun resolution involves not only the core language network but also the network of strategic decision-making. These results coincide with the main assumption of the game-theoretic model: agents are making strategic decisions according to their belief in expected utilities. However, due to the limitation of EEG settings, it is extremely difficult to collect as well as analyze the neural data from both a speaker and a hearer at the same time during a conversation. To perform a closer test to the intersubjectivity of the agents assumed in the game model, the analysis can be further developed in the following two ways. First, to compare the predictions of the current model with corpus study results. A corpus study may provide with sentences in a certain context, and it may help to distinguish a speaker's intention from a hearer's interpretation. Second, to adopt psycholinguistic skills in the design of pretest. The psycholinguistic skills may help to separate the role of the speaker from that of the hearer by setting the pretest and the experiments apart.

References

1. Garvey, C., Caramazza, A.: Factors influencing assignment of pronoun antecedents*. *Cognition* **3**(3), 227–243 (1975)
2. Crawley, R., Stevenson, R., Kleinman, D.: The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* **19**(4), 245–264 (1990)
3. Almor, A.: Noun-phrase anaphors and focus: The informational load hypothesis. *Psychological Review* **106**(4), 748–765 (1999)
4. Garnham, A.: Models of processing: Discourse. *WIREs Cognitive Science* **1**(6), 845–853 (2010)
5. Grice, P.: Logic and conversation. In: Cole, P., Morgan, J. (eds.) *Syntax and Semantics*, pp. 41–58. Academic Press, New York (1975)
6. Lewis, D.: *Convention*. Harvard University Press, Harvard (1969)
7. Parikh, P.: Communication and strategic inference. *Linguistics and Philosophy* **14**, 473–531 (1991)
8. Parikh, P.: *The Use of language*. CSLI Publications, Stanford, CA (2001)
9. Clark, R.: *Meaningful games: Exploring language with game theory*. The MIT Press, Massachusetts (2011)

10. Clark, R., Parikh, P.: Game theory and discourse anaphora. *Journal of Logic, Language and Information* **16**, 265–282 (2007)
11. Frank, M. C., Goodman, N. D.: Predicting pragmatic reasoning in language games. *Science* **336**, 998 (2012)
12. Franke, M.: Signal to act: Game theory in pragmatics. Ph.D. thesis, Institute for Logic, Language and Computation, University of Amsterdam (2009)
13. Franke, M., Jäger, G.: Pragmatic back-and-forth reasoning. In: Reda, S. P. (eds.) *Pragmatics, Semantics, and the Case of Scalar Implicatures*, pp. 170–200. Palgrave, London (2014)
14. Brochhagen, T.: Signalling under uncertainty: Interpretative alignment without a common prior. *The British Journal for the Philosophy of Science*, axx058. <http://doi.org/10.1093/bjps/axx058>
15. Burkhardt, P.: Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language* **98**, 159–168 (2006)
16. Ledoux, K., Gordon, P. C., Camblin, C. C., Swaab, T. Y.: Coreference and lexical repetition: Mechanisms of discourse integration. *Memory and Cognition* **35**(4), 801–815 (2007)
17. Nieuwland, M. S., Petersson, K. M., Van Berkum, J. J. A.: On sense and reference: examining the functional neuroanatomy of referential processing. *NeuroImage* **37**(3), 993–1004 (2007)
18. McMillan, C. T., Clark, R., Gunawardena, D., Ryant, N., Grossman, M.: fMRI evidence for strategic decision-making during resolution of pronoun reference. *Neuropsychologia* **50**(5), 674–687 (2012)
19. Brodbeck, C., Pyllkänen, L.: Language in context: Characterizing the comprehension of referential expressions with MEG. *NeuroImage* **147**, 447–460 (2017)
20. Huettel, S. A., Song, A. W., McCarthy, G.: Decisions under uncertainty: Probabilistic context influences activation of prefrontal and parietal cortices. *Journal of Neuroscience* **25**(13), 3304–3311 (2005)
21. Vickery, T., Jiang, Y.: Inferior parietal lobule supports decision making under uncertainty in humans. *Cerebral Cortex* **19**(4), 916–925 (2009)
22. Levinson, S. C.: *Pragmatics*. Cambridge University Press, New York (1983)
23. Sperber, D., Wilson, D.: *Relevance: Communication and cognition*. 2nd edn. Blackwell, Oxford (1995, 2004)
24. Jäger, G.: Game-theoretical pragmatics. In: van Benthem, J., ter Meulen, A. (eds.) *Handbook of Logic and Language*, pp. 467–491. Elsevier, Amsterdam (2011)
25. van Rooij, R., Franke, M.: Optimality-theoretic and game-theoretic approaches to implicature. *The Stanford Encyclopedia of Philosophy* (2015) <http://plato.stanford.edu/entries/implicatureoptimality-games/>
26. Benz, A., Stevens, J.: Game-theoretic approaches to pragmatics. *Annual Review of Linguistics* **4**, 173–191 (2018)
27. Stickel, C., Fink, J., Holzinger, A.: Enhancing universal access—EEG based learnability assessment. *Lecture Notes in Computer Science LNC* **4556**, 813–822 (2007)
28. Zheng W., Lu B.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development* **7**(3), 162–175 (2015)