

# Understanding Pragmatic Reasoning Through Eye Movement Patterns in Reference Games

Tymur Mykhalievskyi  
7031100

March 2, 2025

## 1 Introduction

If one says “I am going to Munich this week. My mother lives there.”, you will interpret this as meaning they are visiting their mother, even though it is not explicitly stated. This is called an implicature, without explicitly stating something one can still deliver the information. Human communication is full of such implicit constructions. One reason for this may be to save cognitive effort. What rules do people unconsciously follow during communication to make it more efficient?

In 1975 a British philosopher Paul Grice finalized four types of maxims (Grice, 1975). Maxim of Quantity: Provide as much information as required, do not provide more information than required. Maxim of Quality: Be truthful, only say that for which you have adequate evidence. Maxim of Relation: be relevant. Maxim of Manner: avoid ambiguity. Going back to the example with traveling, we can assume a speaker is obeying the maxims. Therefore, the information is relevant and the right amount is provided, so the second sentence about where the mother lives is not just a disconnected fact. Hence, we build an implicature that one is visiting their mother.

One way to study this is through reference games. In these games, participants engage in a collaborative task, often involving the identification or description of objects, where effective communication and reasoning play key roles. Over the years, these reference games have become a popular experimental paradigm to explore how individuals reason about others’ intentions and strategies in communication (Frank & Goodman, 2012; Franke & Degen, 2016). A simple example is presented in Figure 1. Imagine someone is talking to you and uses the word “blue” to refer to one of these objects. Which object are they talking about? If you answered blue square, congrats, it is considered the correct solution. Do not worry, if you are confused. If we consider the possibilities of the speaker, the two completely unambiguous messages available to them are “green” and “circle”. Hence, if they would have referred to one of the other objects, there are clear messages to do

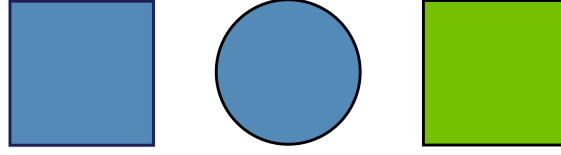


Figure 1: An example of reference game. Same example is shown in Frank and Goodman (2012). A speaker utters "blue", which object are they referring to?

that. Thus, we are left with messages "blue" and "square". Although the message "blue" corresponds to two objects, the blue circle can be referred to unambiguously by using message "circle". Similar logic can be applied to the message "square". So both of them can be inferred to point to the blue square. All this reasoning is built upon the Gricean maxims, as we expect from the speaker to be as concise, unambiguous, relevant and truthful as they can be.

On the other hand, one could notice that the reference games are not as intuitive as the traveling example. It is still a limitation that we will have to keep for now. And this study could also shed light on this problem by understanding what exactly people are doing to solve this kind of problems.

In order to deepen the understanding, a formal model was developed, it is called Rational Speech Act model. It tries to mimic a recursive sequence of reasoning between speaker and listener (Franke & Degen, 2016). Despite significant progress in understanding the cognitive processes underlying these tasks, much remains unknown about the specific strategies individuals employ when solving particular problems within these games.

This study seeks to expand on prior research by incorporating a novel dimension: tracking participants' eye gaze during reference games. Eye gaze offers valuable insight into how people process information, make decisions, and employ strategies. By capturing where and when participants direct their attention, we can gain a deeper understanding of the cognitive mechanisms at play, including how individuals prioritize certain visual cues and how these cues influence their reasoning strategies. This approach has shown to be a very insightful tool (Vigneau, Caissie, & Bors, 2006).

In particular, this study aims to answer the question: How do gaze patterns correlate with the accuracy and strategies used to solve specific communicative challenges in reference games? By integrating eye-tracking data with the analysis of reasoning in these games, this paper contributes to a richer understanding of the decision-making processes involved in collaborative communication and problem-solving.

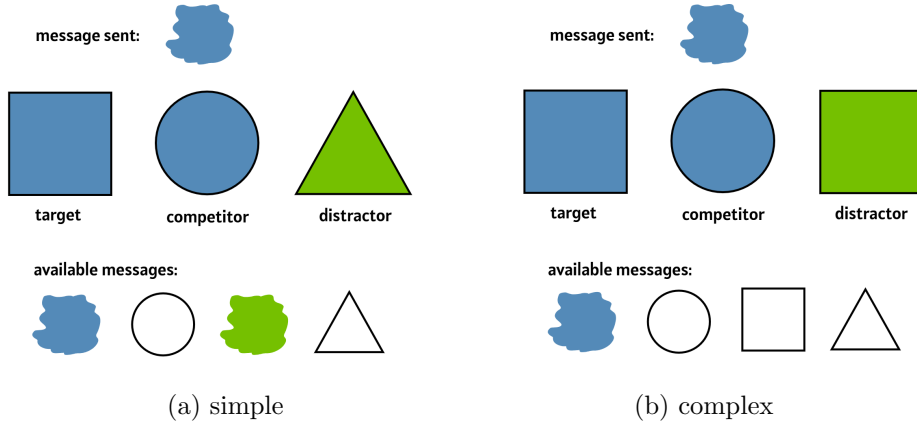


Figure 2: Two instances of reference games with different difficulties.

## 2 Related Work

### 2.1 Reference Games

Although, the idea of communication as a signaling game goes back to Lewis (1969), we will be focusing on a type of a signaling game called a reference game as presented in Frank and Goodman (2012). The game tries to mimic the challenges of communication that people face on daily basis, as we discussed in the section 1. At first, an instance of a reference game looked as in Figure 1, that is, no information about the available messages is given to the speaker and listener. The later instances, on the other hand, include this information. The examples are shown in Figure 2. The goal for newer version is the same, that is, to identify which object a speaker is referring to. This change allows for more controlled experiments by increasing the variability in setups. In addition, it should improve clarity, for instance, participants should not wonder, why wouldn't the speaker just utter the location of the object instead of specifying the properties.

Let's take a closer look at Figure 2a. An uttered message is presented in the top. We will denote the object being referred to as a target, a competitor is an object that shares the message property with the target. While a distractor does not share the sent message property, but could share another property with the target depending on the difficulty of the trial. Note that obviously, captions target, competitor and distractor are not available to the participants. The difference between the simple and complex trials in Figure 2 mainly in how the distractor is constructed. In particular, in the simple trial it does not share any properties with the target, while in the complex it does. The simple example Figure 2a can be solved without considering the distractor. That is, one could count the matching messages from the available ones. In this case, it would be 1 for blue square, 2 for

blue circle and 2 for green triangle. Hence, the target is blue square, as "blue" is the only message that could refer to it. This way of solving is not necessarily what people tend to do, but it is one way of interpreting the difference between simple and complex trials. Because if you apply the same strategy to the complex example in Figure 2b, both target and competitor have two matching messages. On the other hand, if you try to solve these examples yourself, you will probably end up recursively reasoning of what the speaker could have said had they had another target. The simple and complex in this case still appear to their names, you will need a more robust recursion in order to do solve the complex one comparing to the simple one.

## 2.2 Rational Speech Act Model

Studying this phenomena needs a formalized approach. One such model, called the Rational Speech Act (RSA) was developed. It mimics how a speaker and a listener reason about each other. A detailed explanation can be found in the manuscript by Frank, Emilsson, Peloquin, Goodman, and Potts (2016) as well as in the article Franke and Degen (2016). We will go through the main ideas of how listener and speaker interact with each other. Firstly, take a look at the matrix  $M_s$  from the Equation 1. Each columns is a one-hot encoding of an objects, in other words this matrix encodes which objects match the literal meaning of each message, this matrix is constructed for the simple example in Figure 2a.

$$M_s = \begin{matrix} & \blacksquare & \bullet & \blacktriangle \\ \color{blue}{\bullet} & 1 & 1 & 0 \\ \circ & 0 & 1 & 0 \\ \color{green}{\bullet} & 0 & 0 & 1 \\ \triangle & 0 & 0 & 1 \end{matrix} \quad (1)$$

Now let us define a listener matrix, each row shows conditional probabilities for the objects given a message. Accordingly a speaker matrix has columns that depict conditional probabilities of messages given an object.

Subsequently we arrive at a literal listener Equation 2 and speaker Equation 3. Simply put a literal speaker would output one of the matching messages with equal probability for the given target. For example, if green triangle is provided for the speaker, they would refer to it be uttering "green" or "triangle" with equal probability. While literal listener would interpret the ambiguous messages with equal probabilities.

$$L_0(M_s) = L(M_s) = \begin{matrix} & \blacksquare & \bullet & \blacktriangle \\ \color{blue}{\bullet} & 0.5 & 0.5 & 0 \\ \circ & 0 & 1 & 0 \\ \color{green}{\bullet} & 0 & 0 & 1 \\ \triangle & 0 & 0 & 1 \end{matrix} \quad (2)$$

$$S_0(M_s) = S(M_s) = \begin{matrix} & \blacksquare & \bullet & \blacktriangle \\ \bullet & & & \\ \circ & & & \\ \bullet & & & \\ \triangle & & & \end{matrix} \begin{bmatrix} 1 & 0.5 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \\ 0 & 0 & 0.5 \end{bmatrix} \quad (3)$$

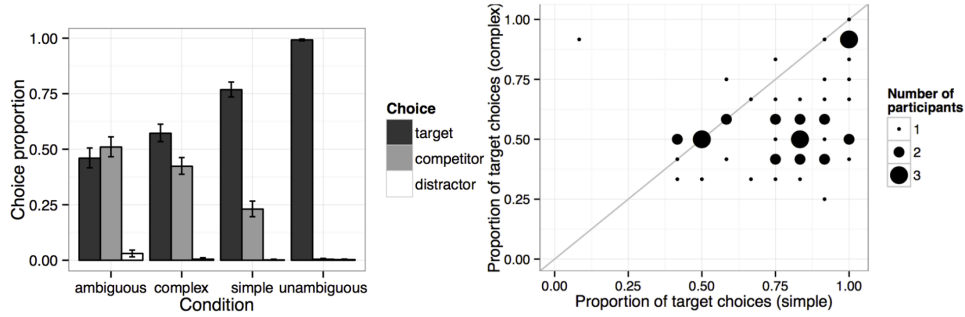
One could see that such approach would not solve even a simple trial. However, if there is a completely unambiguous message, the literal listener would be able to correctly identify the target. The way we derived the two matrices is just a normalization within columns or rows, correspondingly for the listener and speaker. We can keep applying this technique recursively, to find more complex listeners and speakers. That is, a speaker would normalize within columns the matrix previously normalized within rows. In this way we can derive an  $L_1$  listener Equation 4, also called a first-order cooperative listener. Note that Frank et al. (2016) and Franke and Degen (2016) apply different strategies to construct  $L_1$  and  $L_2$  listeners, we will stick to the Franke and Degen (2016) variation.

$$L_1(M_s) = L(S(M_s)) = \begin{matrix} & \blacksquare & \bullet & \blacktriangle \\ \bullet & & & \\ \circ & & & \\ \bullet & & & \\ \triangle & & & \end{matrix} \begin{bmatrix} 0.66 & 0.33 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$L_1$  listener gives the highest probability to the target with message blue. Repeating this procedure further to get to deeper recursion increases the probability of target being chosen. In addition, RSA model has a greed parameter  $\alpha$  which amplifies the probabilities.  $\alpha = \infty$  would result in simply choosing the object with the highest probability. Now let's take a look at the complex case and see how it differs from the simple one. The matrix  $M_c$  is given in Equation 5.

$$M_c = \begin{matrix} & \blacksquare & \bullet & \blacksquare \\ \bullet & & & \\ \circ & & & \\ \square & & & \\ \triangle & & & \end{matrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (5)$$

Going through the same steps to derive the  $L_1$  listener, we get Equation 6.



(a) Proportions of target, competitor and (b) Proportion of target choices in simple and complex conditions by participant. distractor choices in their experiment.

Figure 3: Plots from Franke and Degen (2016).

$$L_1(M_c) = L(S(M_c)) = \begin{matrix} \text{blue square} & \text{blue circle} & \text{green square} \\ \text{blue cloud} & \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 1 & 0 \\ 0.33 & 0 & 0.66 \\ 0 & 0 & 0 \end{bmatrix} \end{matrix} \quad (6)$$

One important difference is that depth of recursion for the  $L_1$  listener is not enough to assign the highest probability to the target. Here the “blue” row has the same probabilities for the distractor and the target. Note that in this case the greed parameter  $\alpha$  would not be able to help. So instead we consider a deeper level of recursion and introduce an  $L_2$  listener Equation 7, also called second-order cooperative listener.

$$L_2(M_c) = L(S(L(M_c))) = \begin{matrix} \text{blue square} & \text{blue circle} & \text{green square} \\ \text{blue cloud} & \begin{bmatrix} 0.6 & 0.4 & 0 \\ 0 & 1 & 0 \\ 0.33 & 0 & 0.66 \\ 0 & 0 & 0 \end{bmatrix} \end{matrix} \quad (7)$$

So as one can see the  $L_2$  can correctly identify the target considering the highest probability. Hence, the main point to take from here is that  $L_1$  listener can solve the simple task, but cannot solve the complex one, while the  $L_2$  listener is able to solve both.

Further expending on this, the previous research shows, that the modeled listeners align with the empirical data. In particular, see Figure 3 taken from Franke and Degen (2016). Figure 3a shows that indeed the difficulty gets harder going from unambiguous to simple and further to complex trials. On the other hand, Figure 3b shows that there are roughly 3 clusters present depending on whether one can solve only simple, both or neither of trials. This strongly supports the alignment with  $L_0$ ,  $L_1$  and  $L_2$  listeners. However,

very important to note that we are only talking about the alignment of RSA model's accuracy with the empirical data, while the concrete strategies are not taken into account.

Now we will proceed further, and make a hypothesis about how people could be solving these problems. A key difference between simple and complex trials is the fact that solving complex trials requires one to consider the distractor as well due to the matching feature with the target, while in the simple one, the distractor can be ignored completely. This can be demonstrated by the following matrix transformations. If one does not take into account the distractor the  $M_s, M_c$  will instead look as in Equation 8 and Equation 9 correspondingly.

$$M'_s = \begin{matrix} & \blacksquare & \bullet \\ \text{■} & \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \\ \text{○} \\ \text{■} \\ \triangle \end{matrix} \quad (8)$$

$$M'_c = \begin{matrix} & \blacksquare & \bullet \\ \text{■} & \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \\ \text{○} \\ \square \\ \triangle \end{matrix} \quad (9)$$

Applying  $L_1$  transformation to the  $M'_s$  we get Equation 10 which accomplishes the same as in Equation 4.

$$L_1(M'_s) = \begin{matrix} & \blacksquare & \bullet \\ \text{■} & \begin{bmatrix} 0.66 & 0.33 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \\ \text{○} \\ \text{■} \\ \triangle \end{matrix} \quad (10)$$

On the other hand, neither  $L_1$  nor  $L_2$  can solve the matrix  $M'_c$ . In fact no depth of recursion is helpful in this case as  $L_0(M'_c) = L_1(M'_c) = L_2(M'_c)$  (Equation 11).

$$L(M'_c) = L(S(M'_c)) = L(S(L(M'_c))) = \begin{matrix} & \blacksquare & \bullet \\ \text{■} & \begin{bmatrix} 0.5 & 0.5 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \\ \text{○} \\ \square \\ \triangle \end{matrix} \quad (11)$$

This leads us to a research question of whether people achieve  $L_1$  accuracy by not considering the distractor and applying reasoning deeper than

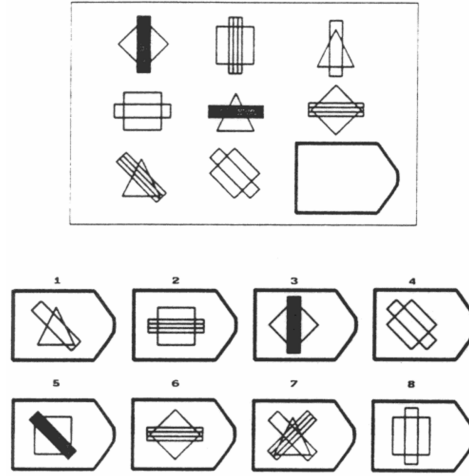


Figure 4: An example of Raven item. The upper part is the matrix, while the bottom is possible answers. The matrix is constructed as follows. The lines orientation is constant within rows. While the shapes and line appearances are obeying the distribution-of-three-values rule. Simply put same three values are present in each row. The correct answer is 5.

one of a literal speaker. Or they include the distractor in their reasoning but simply lack the depth of recursion therefore failing to solve the complex trials.

### 2.3 How Eye Tracking is Useful

We will take a look at a related field with different kind of tasks, this strategy has shown to be particularly informative and insightful there. The Raven Progressive Matrices, commonly referred to as the Raven Tests, are a set of nonverbal intelligence tests designed to measure abstract reasoning and problem-solving abilities through pattern recognition and logical inference. Such test usually contains 8 objects arranged in a 3 by 3 grid with one object missing, as well as the set of possible answers displayed below the matrix. Each matrix either has a particular rule it is constructed by or a mix of them. An example is presented in Figure 4.

Researches suggest that there are two main strategies for solving the Raven Tests constructive matching and response elimination (Bethell-Fox, Lohman, & Snow, 1984) and later followed by Vigneau et al. (2006). The former is described as successively finding rules by which the matrix is constructed, until the answer is fully derived. And the later means that rather than going through the matrix, one goes over the possible answers and eliminates them one by one, ending up with the correct one in the end. The less efficient of these, response elimination, seemed to be used more by lower



ability subjects on more difficult items.

The two strategies can be identified by the patterns of one’s attention, hence, eye gaze. The constructive matching being focused on the matrix and systematically going through rows and columns of it. Carpenter, Just, and Shell (1990) expand on the eye tracking experiments in this research question by recording eye gaze as well as the verbal comments during the process of solving the tasks. A very detailed sequence of actions is acquired therefore giving an insight into how one uses the constructive matching strategy to solve Raven Progressive Matrices. On the other hand, the response elimination involves a lot of toggling between the possible answers and the matrix. In order to deepen the understanding in this problem Vigneau et al. (2006) develop a set of features to encode ones attention. Such features include for example Time on Matrix, Time on Alternatives (possible answers) or Number of Toggles between the possible answers and the matrix. The authors proceed to report the correlation between the features and the percent of ones correct answers. Indeed, the results show statistically significant negative correlation of Time on Alternatives and Number of Toggles with overall score. These findings further supports theory about the difference in effectiveness in the two strategies.

One can see based on these studies why and how the eye tracking is useful in the reference games. In our case as discussed in the end of subsection 2.2 there are multiple ways people could reach  $L_1$  accuracy. Therefore suggesting that the two potential strategies would be distinguished by the use of distractor. At the same time, there are still  $L_0$  and  $L_2$  listeners present in the experiment which makes the distinction more difficult. On the other hand, as there is no previous work on eye tracking reference games, the study is also highly exploratory.

## 2.4 Out of Lab Eye Tracking

Up until now most of the eye tracking studies have been using the laboratory equipment in order to conduct the experiments. This is very important as a reliable and precise method is needed for such experiments. On the other hand, this approach requires people to be physically present in the laboratory, which makes the experiment far more difficult to conduct in comparison to participants answering a series of questions on their laptops. Hence, a different approach was chosen. This experiment will incorporate participants’ webcams to get collect the eye tracking part of the data. In particular a library called WebGazer is used Papoutsaki et al. (2016). Details about the implementation will be discussed in the following sections.

Although, on the first glance, the effectiveness of such approach can be debatable, there is work in favor of the method. Starting from the article Semmelmann and Weigelt (2018) where they take a look into online webcam-based eye tracking comparing it to a respective in-lab experiment.

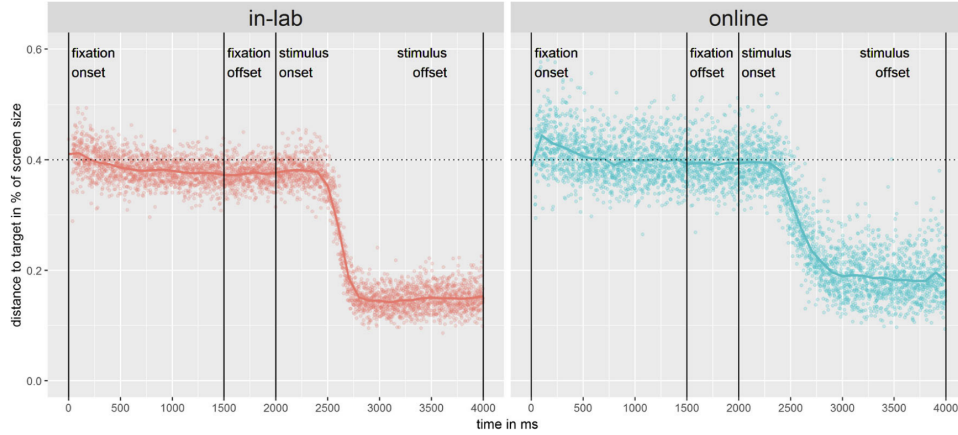


Figure 5: Figure from Semmelmann and Weigelt (2018). Fixation task results. Each dot denotes a single recorded data point in distance to a target in percentage of screen size over time.

Along with a more fresh research article which also makes this comparison (Wisiecka et al., 2022). Both of them conclude that while WebGazer is still inferior to the lab equipment in terms of precision, the measurements are reasonably accurate. In particular, taking a look at the results of Semmelmann and Weigelt (2018) shown in Figure 5. This figure depicts a particular fixation on the target which was shown after 2000 ms. It takes some time for one to react and for the software to capture the eye movement. Then we observe the saccade in both settings, on average the saccade took 450 ms (750 ms in the online case). The accuracy was 171 px (207 px online), which translates to about  $3.94^\circ$  visual angle in the in-lab setting. In addition, it is visible that the online setting has higher variance.

Taking into account the fact that each problem statement by itself consists of multiple objects located on one page, it is not hard to setup them far apart to mitigate the decline in precision. The shorter saccade length is not of the essence in our case.

### 3 Methods

#### 3.1 General information

Participants start the experiment with the general instructions as well as being notified that their consent is obtained. They proceed to the training trials, these invite them to do the speaker’s job. Some trials are unambiguous and some are completely ambiguous. Then participants are to complete the calibration procedure which involves clicking on points on their screen. The results are saved to the model. Then the accuracy of the resulting eye tracker is assessed. If a certain accuracy threshold is not passed, the participants

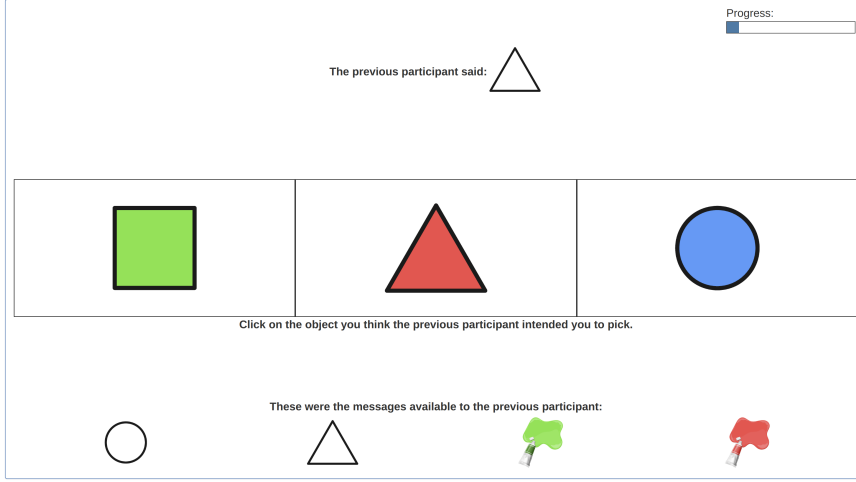


Figure 6: Example of a concrete trial in our experiment.

must redo the calibration procedure. The following is the main part which has 8 simple and 8 complex trials. They look very similar to the ones shown in Figure 2, but they are randomly generated to mitigate any potential biases. Then, there are two more trials, one simple, one complex, each one asks participants about the reason of their answer. This has shown to be useful in the past (Mayn & Demberg, 2023).

### 3.2 Eye Tracking Data

As mentioned before, the eye gaze data is collected via javascript library WebGazer (Papoutsaki et al., 2016). The functionality is nothing but a linear regression model predicting x and y coordinates in pixels. It also provides information about the position and size of the face and eyes on the input video. In addition, the library provides a calibration procedure to adjust the predictions for the particular environment. Thus, no video or audio information is stored at any point in time.

As was described in the subsection 2.4, the layout of the main experiment is made to account for the decrease in accuracy. The example is presented in Figure 6.

### 3.3 Data Analysis

The data analysis is expected to be similar to the one made in Vigneau et al. (2006). In our case the eye tracking features will be: time on available messages, time on message sent, time on distractor, time on target, time on competitor, proportional time on distractor, proportional time on competi-

tor, number of toggles (number of times one shifts between available objects and messages), rate of toggling. In addition, two different target features will be available, they are accuracy on simple and accuracy on complex trials. Firstly a pairwise correlation of features will be assessed. In addition we plan to report fit of linear regressions of multiple features with different targets, or alternatively include the condition as a feature and checking if it is possible to build a linear regression which could predict accuracy on both simple and complex trials depending.

As there is no previous research available on the topic, the data analysis will also include the exploratory part. Although, it might give some wrong intuitions, it is highly important to understand what kind of statistical analysis method could be useful in this kind of problem. This section will include more complicated models as CNNs applied to heatmaps to see if it's possible to predict accuracy based of the heatmap. While this is not very helpful in terms of interpretability, it still could give some intuition in terms whether there even is something to look for. In addition some clustering techniques could be applied.

## 4 After Seminar Work

This proposed research aims to bridge several important gaps in our understanding of pragmatic reasoning through an innovative methodological approach. By combining reference games with eye-tracking technology, we can gain new insights into the cognitive processes underlying pragmatic inference.

Eye-tracking methodology has proven particularly valuable in revealing problem-solving strategies, as demonstrated by research on Raven's Progressive Matrices (Vigneau et al., 2006). The difference between constructive matching and response elimination strategies compared through gaze patterns, suggests that similar insights might be gained in reference games.

The insights gained from this research could have broad implications for our understanding of human communication and reasoning. By revealing the actual strategies people employ during pragmatic inference, we can better understand how communication efficiency is achieved and potentially identify ways to support more effective communication in various contexts.

Furthermore, this work could contribute to the ongoing research Duff, Mayn, and Demberg (2024); Mayn and Demberg (2022); Mayn, Duff, Bila, and Demberg (2024); Mayn, Loy, and Demberg (2024) of pragmatic reasoning by providing empirical evidence of the actual cognitive processes involved. As well as help with further development of more intuitive tasks. This could lead to more refined models that better reflect human reasoning patterns and potentially inform the development of more natural human-computer interaction systems.

## References

- Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence*, 8(3), 205-238. Retrieved from <https://www.sciencedirect.com/science/article/pii/0160289684900096> doi: [https://doi.org/10.1016/0160-2896\(84\)90009-6](https://doi.org/10.1016/0160-2896(84)90009-6)
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review*, 97, 404-431. doi: 10.1037/0033-295x.97.3.404
- Duff, J., Mayn, A., & Demberg, V. (2024). *The role of reinforcement learning in pragmatic reasoning tasks: Modeling individual differences in act-r* [Presentation].
- Frank, M. C., Emilsson, A. G., Peloquin, B., Goodman, N. D., & Potts, C. (2016). *Rational speech act models of pragmatic reasoning in reference games*. PsyArXiv. Retrieved from [osf.io/preprints/psyarxiv/f9y6b](https://osf.io/preprints/psyarxiv/f9y6b) doi: 10.31234/osf.io/f9y6b
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998-998. Retrieved from <https://www.science.org/doi/abs/10.1126/science.1218633> doi: 10.1126/science.1218633
- Franke, M., & Degen, J. (2016). Reasoning in reference games: Individual- vs. population-level probabilistic modeling. *PLOS ONE*, 11(5), 1-25. Retrieved from <https://doi.org/10.1371/journal.pone.0154854> doi: 10.1371/journal.pone.0154854
- Grice, H. P. (1975). Logic and conversation. *Syntax and semantics*, 3, 43-58.
- Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA, USA: Wiley-Blackwell.
- Mayn, A., & Demberg, V. (2022). Individual differences in a pragmatic reference game [Book Section]. In *Proceedings of CogSci 44* (p. 3016-3022).
- Mayn, A., & Demberg, V. (2023, 06). High Performance on a Pragmatic Task May Not Be the Result of Successful Reasoning: On the Importance of Eliciting Participants' Reasoning Strategies. *Open Mind*, 7, 156-178. Retrieved from <https://doi.org/10.1162/opmi.a.00077> doi: 10.1162/opmi.a.00077
- Mayn, A., Duff, J., Bila, N., & Demberg, V. (2024). *Not as rational as rsa predicts: Failure to reason about alternative messages* [Presentation].
- Mayn, A., Loy, J. E., & Demberg, V. (2024). Beliefs about the speaker's reasoning ability influence pragmatic interpretation: Children and adults as speakers.

- Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). Webgazer: Scalable webcam eye tracking using user interactions. In *Proceedings of the 25th international joint conference on artificial intelligence (ijcai)* (pp. 3839–3845).
- Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, 50(2), 451–465. Retrieved from <https://doi.org/10.3758/s13428-017-0913-7> doi: 10.3758/s13428-017-0913-7
- Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence*, 34(3), 261–272. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0160289605001248> doi: <https://doi.org/10.1016/j.intell.2005.11.003>
- Wisiecka, K., Krejtz, K., Krejtz, I., Sromek, D., Cellary, A., Lewandowska, B., & Duchowski, A. (2022). Comparison of webcam and remote eye tracking. In *2022 symposium on eye tracking research and applications*. New York, NY, USA: Association for Computing Machinery. Retrieved from <https://doi.org/10.1145/3517031.3529615> doi: 10.1145/3517031.3529615