Predicting Decision in Human-Agent Negotiation using functional MRI

EunKyung Kim, Sarah I. Gimbel, Aleksandra Litvinova, Jonas T. Kaplan and Morteza Dehghani
{eunkyung; sgimbel; alitvino; jtkaplan; mdehghan}@usc.edu
University of Southern California, Los Angeles, CA 90089 USA

Abstract

The importance of human-agent negotiation, and the role of emotion in such negotiations, have been emphasized in human-agent interaction research. Thus far, studies have focused on behavioral effects, rather than examining the neural underpinnings of different behaviors shown in human-agent interactions. Here, we used a multi-round negotiation platform, instead of the more common single-shot negotiation, and were able to find distinct brain patterns in emotion-related regions of the brain during different types of offers. Using multi-voxel pattern analysis to analyze brain imaging data acquired during functional MRI scanning, we show that it is possible to predict whether the negotiator concedes, does not change, or asks for more during the negotiation. Most importantly, we demonstrate that left dorsal anterior insula, which is known to be an emotion-related brain region, shows a different pattern of activity for each of the three offer types.

Keywords: Decision-making; Negotiation; fMRI

Introduction

We encounter artificial agents frequently in our daily lives. From movie ticket vending machines to rescue robots in a disaster, more and more artificial agents are interacting with people. As interactions between humans and agents become common, effort has been made to make these interactions more effective. For example, relevant lines of research have investigated how to make users more engaged in the interactions with agents. Moreno, Mayer, Spires, and Lester (2001) showed that interactive pedagogical agents who communicate with students via speech promote meaningful learning in multimedia lessons, and agent’s visual presence itself does not provide any cognitive advantage for students’ learning. Castellano, Pereira, Leite, Paiva, and McCowan (2009) demonstrated that social interaction-based features, such as user behavior and contextual information, predict user engagement with a robot companion better than non-verbal behavior features. There has also been research into physiological responses to artificial agents, for example examining heart rate or stress detection during human-agent interaction. Nenonen et al. (2007) showed that a user’s gaming experience was enhanced when the user’s heart rate was used to control an interactive game. Zhai, Barreto, Chin, and Li (2005) also reported that users’ physiological signals, such as pupil diameters, could be used to detect changes in stress level while interacting with a computer. The goal of these studies is to achieve the most effective human-agent interaction, and these approaches have expanded to various branches of human-agent interaction research.

Human-agent negotiation is among the most studied branches of human-agent interaction. Several studies have attempted to simulate human behavior in human-agent negotiation. For example, Byde, Yearworth, Chen, and Bartolini (2003) designed a system that automates price negotiation between buyers and sellers, using a belief function that estimates the distribution of prices and is updated by prior negotiations. The system passed the limited version of the Turing test; human negotiators could not distinguish between a software agent and a human negotiator. Lin, Kraus, Wilkenfeld, and Barry (2008) developed an agent that is capable of negotiating with human counterparts under conditions of incomplete information. Instead of a traditional quantitative decision-making model, a new model that learns information about the counterpart was used to build the agent so that it can achieve better negotiation agreements than human counterparts. These systems have shown the possibility of artificial agents that can reach better negotiation agreements with humans than human negotiators.

In order to make human-agent negotiations more effective, it is important to consider the role of emotions in human-agent interaction. There is now considerable amount of research showing that various emotions expressed by agents can either enhance or hinder interactions (Beale & Creed, 2009; Kim, Dehghani, Kim, Carnevale, & Gratch, 2014). One example of enhanced interactions is demonstrated by Brave, Nass, and Hutchinson (2005), who showed that agents showing emphatic emotion (e.g., showing a happy face when the user won the game and showing a sad face when the user lost the game) resulted in greater positive responses from users compared to when agents showed self-oriented emotion. Also, Van Kleef, De Dreu, and Manstead (2004) showed that people tend to concede more to an angry counterpart than to a happy counterpart.

The majority of previous research work has focused on replicating human-human negotiation results, or achieving better outcomes in human-agent negotiations. We believe that by examining the neural factors underlying behavior, we could advance our understanding about the cognitive processes involved in human-agent negotiations. To examine the neural underpinnings of human-agent negotiations, it is important to understand the functional organization and neurological processes in the human brain involved in such interactions. The use of functional Magnetic Resonance Imaging (fMRI) allows the investigation of the neural substrates of these human-agent negotiations.

Previous fMRI studies have identified specific emotion-related brain regions that play a significant role in social functions: the insular cortex (Damasio et al., 2000; Ruiz et al., 2013) and the amygdala (LeDoux, 2003). Increased insula activation was reported when cooperators see their partner defect (Rilling, Dagenais, Goldsmith, Glenn, & Pagnoni, 2008) or when people are presented with unfair offers (Sanfey,
Rilling, Aronson, Nystrom, & Cohen, 2003). Haruno and Frith (2010) showed that fairness of the negotiation outcome can be predicted based on patterns of activity in amygdala. Activity in these brain regions is believed to affect human behaviors during various types of negotiations, as emotion plays a key role in social interactions that are vital during negotiations (Hess & Bourgeois, 2010). For example, it has been shown that getting an unfair offer or rejecting an offer in the Ultimatum game triggers negative emotions, such as anger (Pillutla & Murnighan, 1996). However, these neurological studies were mostly limited to single-shot negotiations such as the Ultimatum game (Sanfey et al., 2003).

In this study, we build on and extend this line of research by studying the relationship between brain patterns and decision-making during general multi-round human-agent negotiations, which are more similar to real-world negotiations than single-shot negotiations. Investigating brain activity during general multi-round human-agent negotiation is vital to understanding how humans react in interactions with agents. Moreover, investigating the neural systems that are active during different types of interactions would allow us to better understand the underlying cognitive processes responsible for the resultant behavior. As far as we know, our work is the first study that uses brain imaging data to predict the course of general multi-round human-agent negotiations.

We hypothesize that activity in emotion-related brain regions plays a key role in decision-making paradigms during human-agent negotiation. This is supported by previous studies showing that emotion plays an important role in decision-making (Heftvedt & Killian, 1999; Loewenstein, Weber, Hsee, & Welch, 2001). We hypothesize that emotions arising during human-agent interaction are an indicator of what type of decision the person is going to make, potentially predicting one’s negotiation behavior.

The paper is structured as follows: First, we introduce the negotiation task we used in our fMRI experiment. Next, we describe our experimental design and the procedure of the experiment. Then, we explain the methods we used for analyses. Finally, we discuss our results and future work.

**Objects Negotiation Task**

Dehghani, Carnevale, and Gratch (2014) introduced the Objects Negotiation Task, a web-based multi-round negotiation task where a participant and a computer agent can make a proposal in turn to negotiate about the distribution of different types of items. Because this task was originally designed for behavioral studies, some changes were made to optimize the task for use in the fMRI scanner. The version of the Objects Negotiation Task used in this fMRI experiment is composed of six phases (figure 1).

1. The participant proposes an offer.
2. The participant waits for the computer agent to accept or reject the offer.
3. If accepted, negotiation ends. If rejected, the participant waits for the computer agent to propose an offer.
4. The participant reviews the computer agent’s offer for five seconds.
5. The participant decides whether to accept or reject the computer agent’s offer.
6. If accepted, negotiation ends. If rejected, the participant waits for five seconds and then is redirected to the first phase.

Each negotiation can be as long as six rounds. On the sixth round, a message is displayed to the participant to indicate that it is the last round. The participant then can either accept the computer agent’s offer or toss a coin. If a participant chooses to toss a coin, he/she gets two of each item if the coin lands on heads or nothing if the coin lands on tails.

**Experiment**

**Participants**

Ten participants (seven female), recruited from the online campus bulletin board at the University of Southern California, took part in the experiment. Participants’ mean age was 27.5 years (± 4.93) and all participants had normal or corrected to normal vision. Each participant was provided with a written informed consent according to the guidelines of the USC Institutional Review Board. All participants were screened to rule out medication use, head trauma, history of neurological disorders, and other serious medical conditions.

**Procedure**

Participants were instructed to read the hypothetical scenario shown below.

You are a restaurant owner in a small town. There has been a major fire in the market providing the necessary fruits for your restaurant and as a result only a limited number of fruits are available. Because of this you have to split the available fruits with another restaurant owner. You and the other owner value each fruit differently. In order to run your restaurant you need to get as many fruits as possible.

In the task that follows, you will negotiate about how to distribute the fruits between you and the other restaurant owner.

Participants learned the rules of the Objects Negotiation Task and performed a practice task. The practice negotiation task had an identical interface to the task performed in the scanner, but with a different set of negotiation items and payoffs. During the practice negotiation task, participants were provided with a trackball mouse similar to the one used in the scanner in order to get used to operating it and moving items around the task interface. After completing the practice negotiation task, participants were checked with a metal detector to ensure that they were safe to go inside the fMRI scanner.

Participants then performed a total of six negotiations in the scanner, each including up to six rounds. Different sets
of negotiation items were used in each negotiation. After the fMRI scan, the participants received a short survey to acquire background information including gender, age, and handedness. All participants were paid $30 for their participation.

Participants were not told that they would be playing with artificial agents. To simulate playing against other humans, we added randomized delays when the computer agent proposed an offer.

Analysis

In our analyses, we aimed to find the best predictor of participants’ offer types based on their brain activities. To achieve this goal, we labeled our data first, and then performed general linear model (GLM) analysis. Lastly, we used the GLM analysis results as input to multi-voxel pattern analyses (MVPA) using two types of feature selection methods.

Data Labeling

All of the participants’ offers were labeled with one of three categories based on their payoff changes from the previous offer: (1) positively-changed offer, (2) not-changed offer, and (3) negatively-changed offer. Participants’ offers on the first rounds of each negotiation were labeled ‘not-changed offer,’ and their offers on all other rounds were compared with their offer on the previous round based on payoffs and then labeled according to the comparison. For example, if the participant’s payoff in the first round was 20, his/her payoff in the second round was 23, and in the third round 21, the first round would be labeled as a ‘not-changed offer,’ the second round as a ‘positively-changed offer’ (because his/her payoff was increased by 3), and the third round as a ‘negatively-changed offer’ (because his/her payoff was decreased by 2). Here, ‘positively-changed offer’ means that the participant concedes more to the other player, ‘not-changed offer’ means that the participant holds, i.e., he/she retains the offer on the previous round, and ‘negatively-changed offer’ means that the participant asks for more.

Before analyzing the fMRI data, we had to exclude data from three participants. The first exclusion was due to incomplete fMRI scan data. In the second case the participant made less than five offers in each offer category, leading to a too small number of offers in each offer category. In the third case we excluded the data because of the participant’s left-handedness, since more than 70% of left-handed people have different functional brain structure compared to right-handed people such as for language processing (Warrington & Pratt, 1973).

General Linear Model Analysis

To compare brain activity between the offer-making period and the non-offer-making period, we ran a general linear model (GLM) analysis using tools from the FMRIB’s Software Library (FSL) (Smith et al., 2004). Data pre-processing for GLM analysis included following steps. First, all participants’ fMRI data were motion-corrected using FSL’s MCFLIRT tool to fix head motion artifacts during scans. Then, non-brain such as a scalp was removed from the data using FSL’s Brain Extraction Tool. Next, spatial smoothing using a 5mm full width at half maximum Gaussian kernel was applied to increase statistical power by improving the signal to noise ratio. Also, slice timing correction for interleaved acquisitions was used to compensate for timing difference between slices of functional images. Finally, high-pass temporal filtering was performed to let high frequencies (containing activities relevant to decision-making) pass and to remove low frequencies such as signal drifts.

After completing data pre-processing, we modeled brain activity during offer-making with a double gamma hemodynamic response function. Brain activity during all other time points were considered baseline.
Multi-Voxel Pattern Analysis

To analyze brain patterns in decision-making during negotiations, we used multi-voxel pattern analysis (MVPA) (Norman, Polyn, Detre, & Haxby, 2006), a machine learning approach for investigating patterns of brain voxels. Instead of analyzing each voxel separately, MVPA takes multiple voxels into account together. This is useful because activity in one voxel cannot be separated from neighboring voxels.

We used GLM analysis results as input to MVPA. As data pre-processing steps for MVPA, we de-trended the data to remove any bias resulting from scanner drift over the acquisition time. Then, we converted the data to z-scores to normalize the range of each voxel. We used leave-one-participant-out cross validation for MVPA, in which a classifier is trained on six participants’ data and then tested with the last participant’s data. We repeated this seven times, leaving each participant out once, then averaged the results to calculate prediction accuracy. A balancer was also included to keep the chance level the same (33%) throughout our analyses because every participant has a different set of offers. Since a balancer chooses a new set of offers in each category whenever it runs, we ran MVPA five times and averaged the results.

Lastly, we applied feature selection methods for choosing the voxels used in our analysis. Feature selection is a common approach to reduce the number of features (voxels) by selecting only relevant features as input to a classifier. Classification performance improves with feature selection because it picks features that vary significantly between categories (Guyon & Elisseeff, 2003). To validate our hypothesis that brain activities on emotion-related regions are closely related to decision-making during negotiations, we used two different feature selection methods for MVPA. For both approaches, we used a linear Support Vector Machine (SVM) classifier to perform classification. In the following sections, we explain each the two feature selection method we used for classifier to perform classification. In the following sections, different feature selection methods for MVPA. For both approaches, we used a linear Support Vector Machine (SVM) classifier to perform classification. In the following sections, we explain each the two feature selection method we used for classifier to perform classification.

Region of Interest Analysis: Insular Cortex

In the region of interest (ROI) analysis, we attempted to predict the offer categories based only on the voxels in the insular cortex, which is known as an emotion-related brain region. The insular cortex on each side of the brain can be divided into three subregions with distinct patterns of connectivity: dorsal anterior insula, connected with dorsal anterior cingulate cortex; ventral anterior insula, connected with pregenual anterior cingulate cortex; and posterior insula, connected with primary and secondary somatomotor cortices (Deen, Pitskel, & Pelphrey, 2011).

We trained the classifier using voxels from each of the six regions separately with feature selection. In our analyses, we used GLM analysis results to compute F-score per each voxel, and then used an analysis of variance measure to select the top 5% of features with the highest F-scores. Each participant’s brain was first transformed into standard MNI space (Evans et al., 1993) to minimize differences from individual brains. After performing this process for all participants, individual-level analyses were combined for a group-level analysis.

Searchlight as a Feature Selection Method

In the searchlight analysis (Kriegeskorte, Goebel, & Bandettini, 2006), a map of classification accuracies is generated by measuring the information in small spheres (radius = 5 voxels) centered on every voxel in brain. As activities on one voxel are inevitably influenced by activities on neighboring voxels, searchlight analysis is a preferable approach to capture local spatial areas that show significantly different activity patterns on each experimental condition. To confirm the strong correlation between brain activities on emotion-related regions and negotiation offer types, and to find other brain regions that show notable pattern differences on each offer type, we used searchlight analysis as a feature selection method for MVPA.

Steps for using searchlight analysis as a feature selection method are as follows. We first generated searchlight maps from each participant’s brain. Then we transformed individual searchlight maps into the standard MNI space, and merged six of them to generate seven merged searchlight maps in total. Next, we generated t-maps using t-tests across six participants versus chance, and thresholded the top 5% of the t-maps and binarized them. After that, we transformed all participants’ functional images (EPIs) into individual EPI space and ran MVPA using thresholded and binarized t-maps as masks for each participant. Finally, we averaged MVPA results to calculate overall prediction accuracy.

RESULTS

ROI MVPA: Left Dorsal Anterior Insula

We hypothesized that activity in the insular cortex would be highly correlated with different types of offers. We performed ROI MVPA with each of the six insula regions (left/ right ventral anterior insula, left/right dorsal anterior insula, and left/right posterior insula), and found that left dorsal anterior insula to be the best predictor of offer types (positively-changed offer, not-changed offer, or negatively-changed offer). The prediction accuracy of ROI MVPA using voxels from left dorsal anterior insula is 43.88%, with a standard error 1.30%. Interestingly, insular regions other than left dorsal anterior insula show chance level performance.

A binomial test shows that the performance of ROI MVPA with left dorsal anterior insula is significantly higher than chance level ($p = 0.0058$), indicating that activity in left dorsal anterior insula predicts offer types in negotiations.

MVPA with Searchlight as a Feature Selection Method

To map the spatial distribution of information, we performed MVPA with searchlight as a feature selection method. Figure 2 shows the overlaid accuracy map for all seven participants. Before overlaying accuracy maps, we thresholded the top 5% of seven t-maps each across six participants and binarized them. Consistent with the finding from the ROI analysis, we find that left dorsal anterior insula is included in six t-maps.
The prediction accuracy for MVPA with searchlight as a feature selection method is 47.45% with a standard error 0.017%. Compared to ROI MVPA, MVPA with searchlight as a feature selection method has the prediction accuracy improved by 3.57% and has much smaller standard error. Two sample t-test results revealed that there is a significant difference in prediction accuracy between MVPA with searchlight as a feature selection method and ROI MVPA with left dorsal anterior insula, \( t(23.124) = 1.6272, p = 0.058 \).

**DISCUSSION**

In this experiment, we investigated the neural correlates of decision-making during negotiations. We classified negotiation offers into three categories: conceding, holding, or asking for more. We then analyzed brain imaging data by offer category to see if we could predict the category of negotiation based on the brain activation. We used functional MRI to capture participants’ brain activity during negotiations, and used MVPA to analyze the fMRI data. ROI MVPA and MVPA with searchlight as feature selection methods were used and both methods resulted in significantly better prediction accuracies than the chance level of 33%. Most notably, MVPA with searchlight as a feature selection method yielded the higher prediction accuracy of 47.45%, indicating the importance of analyzing neighboring voxel clusters together.

Our results reveal that there are distinct brain patterns across participants for each type of offer. More specifically, activity in left dorsal anterior insula, which is a well-known emotion-related brain region, was found to play a key role in distinguishing offer types. This is in line with our hypothesis that activations in emotion-related brain regions would be closely related to decision-making. Emotions provoked while the participant is interacting with the computer agent during negotiations mediate the participant’s negotiation behaviors, and these processes are captured in fMRI data as a form of increased blood flow in emotion-related brain regions. Thus, this confirms not only the importance of the role of emotion in human-agent interaction, but also the possibility of interpreting the underlying processes during negotiations with fMRI data.

Furthermore, the results indicate that we can predict negotiation offers based on brain activities. Offer type prediction results from MVPA with two types of feature selection methods support the feasibility of successful predictions of negotiation behaviors. This has implications for human-agent negotiation research, allowing us to perform more detailed predictions of negotiation behavior based on fMRI data as compared to predictions based on behavioral research. For example, we can conduct an fMRI study on how engaged participants are during negotiations under various conditions. In addition to non-verbal features that were found to be related to user engagement by Castellano et al. (2009), introducing brain imaging data would give us extra information on degree of user engagement. With these approaches, we can come close to achieving the most effective human-agent interaction.

Also, our results suggest that there are more brain regions correlated with negotiation offer types. For instance, right posterior supramarginal gyrus and right hippocampus were included in all t-maps that were generated with searchlight maps. Generally, decision-making during negotiations are affected by one’s empathy towards others (Loewenstein & Lerner, 2003), which is related to brain activities on right supramarginal gyrus, and remembering previous offers to update one’s knowledge (Zeng & Sycara, 1998), which is related to brain activities on right hippocampus. Thus, these findings would allow us to track down more details of decision-making during negotiations.

In this study, we looked at only one previous round of negotiation, but it is possible that negotiators actually take into account the whole history of negotiation when proposing the next offer. In that case, studying a longer history would result in better prediction of negotiators’ offers. As a first approach studying general multi-round human-agent negotiations, we were able to show that we can do offer type prediction during negotiations with one previous round using fMRI data. In future studies, we plan to take longer history into consideration. Also, we plan to replicate our results using larger sample sizes. Even though we had only seven participants in this study, we expect to see the same results in a large number of participants because we were able to do a significantly better offer type prediction than the chance level with large effect sizes. In addition, we would like to note that the probability of replication is dependent not on the sample size but on the \( p \) value (Killeen, 2005).

In conclusion, we examined the relationship between brain patterns and types of negotiation offers. Unlike previous fMRI studies that were limited to single-shot negotiations, we studied general multi-round human-agent negotiations. Our results demonstrate that we can predict behaviors from multi-round negotiations using fMRI data. We believe this finding could help enrich human-agent negotiation, and hope to find more work on this topic.
Acknowledgments

This research is supported by AFOSR FA9550-14-1-0364.

References


