Persuasion for Good:
Towards a Personalized Persuasive Dialogue System for Social Good

Xuewei Wang[*1], **Weiyan Shi**[*2], **Richard Kim**[2], **Yoojung Oh**[2]
Sijia Yang[3], **Jingwen Zhang**[2] **and Zhou Yu**[2]

[1] Zhejiang University,[2] University of California, Davis,[3] University of Pennsylvania

cheriewang@zju.edu.cn,{wyshi, khgkim, yjeoh}@ucdavis.edu,
sijia.yang@asc.upenn.edu,{jwzzhang, joyu}@ucdavis.edu

Abstract

Developing intelligent persuasive conversational agents to change people's opinions and actions for social good is the frontier in advancing the ethical development of automated dialogue systems. To do so, the first step is to understand the intricate organization of strategic disclosures and appeals employed in human persuasion conversations. We designed an online persuasion task where one participant was asked to persuade the other to donate to a specific charity. We collected a large dataset with 1,017 dialogues and annotated emerging persuasion strategies from a subset. Based on the annotation, we built a baseline classifier with context information and sentence-level features to predict the 10 persuasion strategies used in the corpus. Furthermore, to develop an understanding of personalized persuasion processes, we analyzed the relationships between individuals' demographic and psychological backgrounds including personality, morality, value systems, and their willingness for donation. Then, we analyzed which types of persuasion strategies led to a greater amount of donation depending on the individuals' personal backgrounds. This work lays the ground for developing a personalized persuasive dialogue system. [1]

1 Introduction

Persuasion aims to use conversational and messaging strategies to change one specific person's attitude or behavior. Moreover, personalized persuasion combines both strategies and user information related to the outcome of interest to achieve better persuasion results (Kreuter et al., 1999; Rimer and Kreuter, 2006). Simply put, the goal of personalized persuasion is to produce desired changes by making the information personally relevant and appealing. However, two questions about personalized persuasion still remain unexplored. First, we concern about how personal information would affect persuasion outcomes. Second, we question about what strategies are more effective considering different user backgrounds and personalities.

The past few years have witnessed the rapid development of conversational agents. The primary goal of these agents is to facilitate task-completion and human-engagement in practical contexts (Luger and Sellen, 2016; Bickmore et al., 2016; Graesser et al., 2014; Yu et al., 2016b). While persuasive technologies for behavior change have successfully leveraged other system features such as providing simulated experiences and behavior reminders (Orji and Moffatt, 2018; Fogg, 2002), the development of automated persuasive agents remains lagged due to the lack of synergy between the social scientific research on persuasion and the computational development of conversational systems.

In this work, we introduced the foundation work on building an automatic personalized persuasive dialogue system. We first collected 1,017 human-human persuasion conversations (PERSUASION-FORGOOD) that involved real incentives to participants. Then we designed a persuasion strategy annotation scheme and annotated a subset of the collected conversations. In addition, we came to classify 10 different persuasion strategies using Recurrent-CNN with sentence-level features and dialogue context information. We also analyzed the relations among participants' demographic backgrounds, personality traits, value systems, and their donation behaviors. Lastly, we analyzed what types of persuasion strategies worked more effectively for what types of personal backgrounds. These insights will serve as important el-

---

ements during our design of the personalized persuasive dialogue systems in the next phase.

2 Related Work

In social psychology, the rationale for personalized persuasion comes from the Elaboration Likelihood Model (ELM) theory (Petty and Cacioppo, 1986). It argues that people are more likely to engage with persuasive messages when they have the motivation and ability to process the information. The core assumption is that persuasive messages need to be associated with the ways different individuals perceive and think about the world. Hence, personalized persuasion is not simply capitalizing on using superficial personal information such as name and title in the communication; rather, it requires a certain degree of understanding of the individual to craft unique messages that can enhance his or her motivation to process and comply with the persuasive requests (Kreuter et al., 1999; Rimer and Kreuter, 2006; Dijkstra, 2008).

There has been an increasing interest in persuasion detection and prediction recently. Hidey et al. (2017) presented a two-tiered annotation scheme to differentiate claims and premises, and different persuasion strategies in each of them in an online persuasive forum (Tan et al., 2016). Hidey and McKeown (2018) proposed to predict persuasiveness by modelling argument sequence in social media and showed promising results. Yang et al. (2019) proposed a hierarchical neural network model to identify persuasion strategies in a semi-supervised fashion. Inspired by these prior work in online forums, we present a persuasion dialogue dataset with user demographic and psychological attributes, and study personalized persuasion in a conversational setting.

In the past few years, personalized dialogue systems have come to people's attention because user-targeted personalized dialogue system is able to achieve better user engagement (Yu et al., 2016a). For instance, Shi and Yu (2018) exploited user sentiment information to make dialogue agent more user-adaptive and effective. But how to get access to user personal information is a limiting factor in personalized dialogue system design. Zhang et al. (2018) introduced a human-human chit-chat dataset with a set of 1K+ personas. In this dataset, each participant was randomly assigned a persona that consists of a few descriptive sentences. However, the brief description of user persona lacks quantitative analysis of users' sociodemographic backgrounds and psychological characteristics, and therefore is not sufficient for interaction effect analysis between personalities and dialogue policy preference.

Recent research has advanced the dialogue system design on certain negotiation tasks such as bargain on goods (He et al., 2018; Lewis et al., 2017). The difference between negotiation and persuasion lies in their ultimate goal. Negotiation strives to reach an agreement from both sides, while persuasion aims to change one specific person's attitude and decision. Lewis et al. (2017) applied end-to-end neural models with self-play reinforcement learning to learn better negotiation strategies. In order to achieve different negotiation goals, He et al. (2018) decoupled the dialogue act and language generation which helped control the strategy with more flexibility. Our work is different in that we focus on the domain of persuasion and personalized persuasion procedure.

Traditional persuasive dialogue systems have been applied in different fields, such as law (Gordon, 1993), car sales (André et al., 2000), intelligent tutoring (Yuan et al., 2008). However, most of them overlooked the power of personalized design and didn't leverage deep learning techniques. Recently, Lukin et al. (2017) considered personality traits in single-turn persuasion dialogues on social and political issues. They found that personality factors can affect belief change, with conscientious, open and agreeable people being more convinced by emotional arguments. However, it's difficult to utilize such a single-turn dataset in the design of multi-turn dialogue systems.

3 Data Collection

We designed an online persuasion task to collect emerging persuasion strategies from human-human conversations on the Amazon Mechanical Turk platform (AMT). We utilized ParlAI (Miller et al., 2017), a python-based platform that enables dialogue AI research, to assist the data collection. We picked *Save the Children*[2] as the charity to donate to, because it is one of the most well-known charity organizations around the world.

Our task consisted of four parts, a pre-task survey, a persuasion dialogue, a donation confirmation and a post-task survey. Before the conversation began, we asked the participants to com-

---

[2] https://www.savethechildren.org/

Role	Utterance	Annotation
ER	Hello, are you interested in protection of rights of children?	Source-related inquiry
EE	Yes, definitely. What do you have in mind?	
ER	There is an organisation called Save the Children and donations are essential to ensure children's rights to health, education and safety.	Credibility appeal
EE	Is this the same group where people used to "sponsor" a child?	
ER	Here is their website, https://www.savethechildren.org/.	Credibility appeal
	They help children all around the world.	Credibility appeal
	For instance, millions of Syrian children have grown up facing the daily threat of violence.	Emotion appeal
	In the first two months of 2018 alone, 1,000 children were reportedly killed or injured in intensifying violence.	Emotion appeal
EE	I can't imagine how terrible it must be for a child to grow up inside a war zone.	
ER	As you mentioned, this organisation has different programs, and one of them is to "sponsor" child.	Credibility appeal
	You choose the location.	Credibility appeal
EE	Are you connected with the NGO yourself?	
ER	No, but i want to donate some amount from this survey.	Self-modeling
	Research team will send money to this organisation.	Donation information
EE	That sounds great. Does it come from our reward/bonuses?	
ER	Yes, the amount you want to donate is deducted from your reward.	Donation information
EE	What do you have in mind?	
ER	I know that my small donation is not enough, so i am asking you to also donate some small percentage from reward.	Proposition of donation
EE	I am willing to match your donation.	
ER	Well, if you go for full 0.30 i will have no moral right to donate less.	Self-modeling
EE	That is kind of you. My husband and I have a small NGO in Mindanao, Philippines, and it is amazing what a little bit of money can do to make things better.	
ER	Agree, small amount of money can mean a lot for people in third world countries.	Foot-in-the-door
	So agreed? We donate full reward each??	Donation confirmation
EE	Yes, let's donate $0.30 each. That's a whole lot of rice and flour. Or a whole lot of bandages.	

Table 1: An example persuasion dialogue. ER and EE refer to the persuader and the persuadee respectively.

plete a pre-task survey to assess their psychological profile variables. There were four sub-questionnaires in our survey, the Big-Five personality traits (Goldberg, 1992) (25 questions), the Moral Foundations endorsement (Graham et al., 2011) (23 questions), the Schwartz Portrait Value (10 questions) (Cieciuch and Davidov, 2012), and the Decision-Making style (4 questions) (Hamilton and Mohammed, 2016). From the pre-task survey, we obtained a 23-dimension psychological feature vector where each element is the score of one characteristic, such as extrovert and agreeable.

Next, we randomly assigned the roles of persuader and persuadee to the two participants. The random assignment helped to eliminate the correlation between the persuader's persuasion strategies and the targeted persuadee's characteristics. In this task, the persuader needed to persuade the persuadee to donate part of his/her task earning to the charity, and the persuader could also choose to donate. Please refer to Fig. 6 and 7 in Appendix for the data collection interface. For persuaders, we provided them with tips on different persuasion strategies along with some example sentences. For persuadees, they only knew they would talk about a specific charity in the conversation. Participants were encouraged to continue the conversation until an agreement was reached. Each participant was required to complete at least 10 conversational turns and multiple sentences in one turn were allowed. An example dialogue is shown in Table 1.

After completing the conversation, both the per-

Dataset Statistics		
# Dialogues		1,017
# Annotated Dialogues (ANNSET)		300
# Participants		1,285
Avg. donation		$0.35
Avg. turns per dialogue		10.43
Avg. words per utterance		19.36
Total unique tokens		8,141
Participants Statistics		
Metric	*Persuader*	*Persuadee*
Avg. words per utterance	22.96	15.65
Donated	424 (42%)	545 (54%)
Not donated	593 (58%)	472 (46%)

Table 2: Statistics of PERSUASIONFORGOOD

suader and the persuadee were asked to input the intended donation amount privately though a text box. The max amount of donation was the task payment. After the conversation ended, all participants were required to finish a post-survey assessing their sociodemographic backgrounds such as age and income. We also included several questions about their engagement in this conversation.

The data collection process lasted for two months and the statistics of the collected dataset named PERSUASIONFORGOOD are presented in Table 2. We observed that on average persuaders chose to say longer utterances than persuadees (22.96 tokens compared to 15.65 tokens). During the data collection phase, we were glad to receive some positive comments from the workers. Some mentioned that it was one of the most meaningful tasks they had ever done on the AMT, which

shows an acknowledgement to our task design.

4 Annotation

Category	Amount
Logical appeal	325
Emotion appeal	237
Credibility appeal	779
Foot-in-the-door	134
Self-modeling	150
Personal story	91
Donation information	362
Source-related inquiry	167
Task-related inquiry	180
Personal-related inquiry	151
Non-strategy dialogue acts	1737
Total	4313

Table 3: Statistics of persuasion strategies in ANNSET.

After the data collection, we designed an annotation scheme to annotate different persuasion strategies persuaders used. Content analysis method (Krippendorff, 2004) was employed to create the annotation scheme. Since our data was from typing conversation and the task was rather complicated, we observed that half of the conversation turns contained more than two sentences with different semantic meanings. So we chose to annotate each complete sentence instead of the whole conversation turn.

We also designed a dialogue act annotation scheme for persuadee's utterances, shown in Table 6 in Appendix, to capture persuadee's general conversation behaviors. We also recorded if the persuadee agreed to donate, and the intended donation amount mentioned in the conversation.

We developed both persuader and persuadee's annotation schemes using theories of persuasion and a preliminary examination of 10 random conversation samples. Four research assistants independently coded 10 conversations, discussed disagreement, and revised the scheme accordingly. The four coders conducted two iterations of coding exercises on five additional conversations and reached an inter-coder reliability of Krippendorff's alpha of above 0.70 for all categories. Once the scheme was finalized, each coder separately coded the rest of the conversations. We named the 300 annotated conversations as the ANNSET.

Annotations for persuaders' utterances included diverse argument strategies and task-related non-persuasive dialogue acts. Specifically, we identified 10 persuasion strategy categories that can be divided into two types, 1) **persuasive appeal** and 2) **persuasive inquiry**. Non-persuasive dialogue acts included general ones such as greeting, and task-specific ones such as donation proposition and confirmation. Please refer to Table 7 in Appendix for the persuader dialogue act scheme.

The seven strategies below belong to **persuasive appeal**, which tries to change people's attitudes and decisions through different psychological mechanisms.

Logical appeal refers to the use of reasoning and evidence to convince others. For instance, a persuader can convince a persuadee that the donation will make a tangible positive impact for children using reasons and facts.

Emotion appeal refers to the elicitation of specific emotions to influence others. Specifically, we identified four emotional appeals: 1) telling stories to involve participants, 2) eliciting empathy, 3) eliciting anger, and 4) eliciting the feeling of guilt. (Hibbert et al., 2007).

Credibility appeal refers to the uses of credentials and citing organizational impacts to establish credibility and earn the persuadee's trust. The information usually comes from an objective source (*e.g.*, the organization's website or other well-established websites).

Foot-in-the-door refers to the strategy of starting with small donation requests to facilitate compliance followed by larger requests (Scott, 1977). For instance, a persuader first asks for a smaller donation and extends the request to a larger amount after the persuadee shows intention to donate.

Self-modeling refers to the strategy where the persuader first indicates his or her own intention to donate and chooses to act as a role model for the persuadee to follow.

Personal story refers to the strategy of using narrative exemplars to illustrate someone's donation experiences or the beneficiaries' positive outcomes, which can motivate others to follow the actions.

Donation information refers to providing specific information about the donation task, such as the donation procedure, donation range, etc. By providing detailed action guidance, this strategy can enhance the persuadee's self-efficacy and facilitates behavior compliance.

The three strategies below belong to **persuasive**

inquiry, which tries to facilitate more personalized persuasive appeals and to establish better interpersonal relationships by asking questions.

Source-related inquiry asks if the persuadee is aware of the organization (i.e., the source in our specific donation task).

Task-related inquiry asks about the persuadee's opinion and expectation related to the task, such as their interests in knowing more about the organization.

Personal-related inquiry asks about the persuadee's previous personal experiences relevant to charity donation.

The statistics of the ANNSET are shown in Table 3, where we listed the number of times each persuasion strategy appears. Most of the further studies are on the ANNSET. Example sentences for each persuasion strategy are shown in Table 4.

We first explored the distribution of different strategies across conversation turns. We present the number of different persuasion strategies at different conversation turn positions in Fig. 1 (for persuasive appeal) and Fig. 2 (for persuasive inquiry). As shown in Fig. 1, *Credibility appeal* occurred more at the beginning of the conversations. In contrast, *Donation information* occurred more in the latter part of the conversations. *Logical appeal* and *Emotion appeal* share a similar distribution and also frequently appeared in the middle of the conversations. The rest of the strategies, *Personal story*, *Self-modeling* and *Foot-in-the-door*, are spread out more evenly across the conversations, compared with the other strategies. For persuasive inquiries in Fig. 2, *Source-related inquiry* mainly appeared in the first three turns, and the other two kinds of inquiries have a similar distribution.
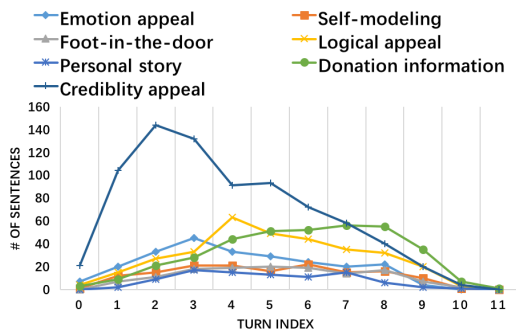

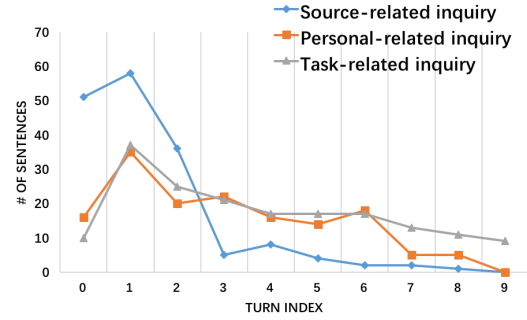Figure 1: Distributions of the seven persuasive appeals across turns.

Figure 2: Distributions of the three persuasive inquiries across turns.
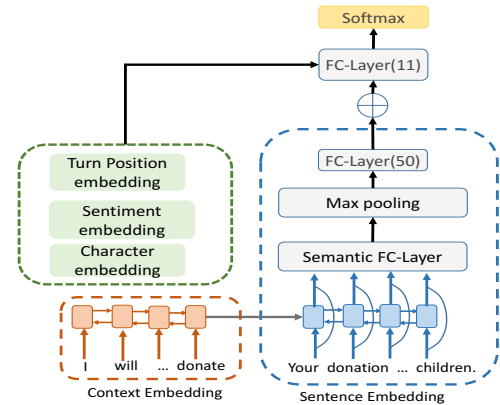
5   Donation Strategy Classification

Figure 3: The hybrid RCNN model combines sentence embedding, context embedding and sentence-level features. "+" represents vector concatenation. The blue dotted box shows the sentence embedding part. The orange dotted box shows the context embedding part. The green dotted box shows the sentence-level features.

In order to build a persuasive dialogue system, we need to first understand human persuasion patterns and differentiate various persuasion strategies. Therefore, we designed a classifier for the 10 persuasion strategies plus one additional "non-strategy" class for all the non-strategy dialogue acts in the ANNSET. We proposed a hybrid RCNN model which combined the following features, 1) sentence embedding, 2) context embedding and 3) sentence-level feature, for the classification. The model structure is shown in Fig. 3.

Sentence embedding used recurrent convolutional neural network (RCNN), which combined CNN and RNN to extract both the global and local semantics, and the recurrent structure may reduce noise compared to the window-based neural network (Lai et al., 2015). We concatenated the word

Persuasion Strategy	Example
Logical appeal	*Your donation could possible go to this problem and help many young children.* *You should feel proud of the decision you have made today.*
Emotion appeal	*Millions of children in Syria grow up facing the daily threat of violence.* *This should make you mad and want to help.*
Credibility appeal	*And the charity is highly rated with many positive rewards.* *You can find reports associated with the financial information by visiting this link.*
Foot-in-the-door	*And sometimes even a small help is a lot, thinking many others will do the same.* *By people like you, making a a donation of just $1 a day, you can feed a child for a month.*
Self-modeling	*I will donate to Save the Children myself.* *I will match your donation.*
Personal story	*I like to give a little money to charity each month.* *My brother and I replaced birthday gifts with charity donations a few years ago.*
Donation information	*Your donation will be directly deducted from your task payment.* *The research team will collect all donations and send it to Save the Children.*
Source-related inquiry	*Have you heard of Save the Children?* *Are you familiar with the organization?*
Task-related inquiry	*Do you want to know the organization more?* *What do you think of the charity?*
Personal-related inquiry	*Do you have kids?* *Have you donated to charity before?*

Table 4: Example sentences for the 10 persuasion strategies.

embedding and the hidden state of the LSTM as the sentence embedding s_t. Next, a linear semantic transformation was applied on s_t to obtain the input to a max-pooling layer. Finally, the pooling layer was used to capture the effective information throughout the entire sentence.

Context embedding was composed of the previous persuadee's utterance. Considering the relatively long context, we used the last hidden state of the context LSTM as the initial hidden state of the RCNN. We also experimented with other methods to extract context and will detail them in Section 6.

We also designed three **sentence-level features** to capture meta information other than embeddings. We describe them below.

Turn position embedding. According to the previous analysis, different strategies have different distributions across conversation turns, so the turn position may help the strategy classification. We condensed the turn position information into a 10-dimension embedding vector.

Sentiment. We also extracted sentiment features for each sentence using VADER (Gilbert, 2014), a rule-based sentiment analyzer. It generates negative, positive, neutral scores from zero to one. It is interesting to note that for *Emotion appeal*, the average negative sentiment score is 0.22, higher than the average positive sentiment score, 0.10. It seems negative sentiment words are used more frequently in *Emotion appeal* because persuaders tend to describe sad facts to arouse empathy in *Emotion appeal*. In contrast, positive words are used more frequently in *Logical appeal*, because persuaders tend to describe more positive results from donation when using *Logical appeal*.

Character embedding. For short text, character level features can be helpful. Bothe et al. (2018) utilized character embedding to improve the dialogue act classification accuracy. Following Bothe et al. (2018), we chose the pre-trained multiplicative LSTM (mLSTM) network on 80 million Amazon product reviews to extract 4096-dimension character-level features (Radford et al., 2017)[3]. Given the output character embedding, we applied a linear transformation layer with output size 50 to obtain the final character embedding.

6 Experiments

Because human-human typing conversations are complex, one sentence may belong to multiple strategy categories; out of the concern for model simplicity, we chose to predict the most salient strategy for each sentence. Table 3 shows the dataset is highly imbalanced, so we used the macro F1 as the evaluation metric, in addition to accuracy. We conducted five-fold cross validation, and used the average scores across folds to compare the performance of different models. We set the initial learning rate to be 0.001 and applied exponential decay every 100 steps. The training batch size was 32 and all models were trained for 20 epochs. In addition, dropout (Srivastava et al.,

---

[3]https://github.com/openai/generating-reviews-discovering-sentiment

2014) with a probability of 0.5 was applied to reduce over-fitting. We adopted the 300-dimension pre-trained FastText (Bojanowski et al., 2017) as word embedding. The RCNN model used a single-layer bidirectional LSTM with a hidden size of 200. We describe two baseline models below for comparison.

Self-attention BLSTM (BLSTM) only considers a single-layer bidirectional LSTM with self-attention mechanism. After finetuning, we set the attention dimension to be 150.

Convolutional neural network (CNN) uses multiple convolution kernels to extract textual features. A softmax layer was applied in the end to generate the probability for each category. The hyperparameters in the original implementation (Kim, 2014) were used.

6.1 Experimental Results

Models	Accuracy	Macro F1
Majority vote	18.1%	5.21%
BLSTM + All features	73.4%	57.1%
CNN + All features	73.5%	58.0%
Hybrid RCNN with different features		
Sentence only	74.3%	59.0%
Sentence + Context CNN	72.5%	54.5%
Sentence + Context Mean	74.0%	58.5%
Sentence + Context RNN	74.4%	59.3%
Sentence + Context tf-idf	73.5%	57.6%
Sentence + Turn position	73.8%	59.4%
Sentence + Sentiment	73.6%	59.7%
Sentence + Character	74.5%	59.3%
All features	**74.8%**	**59.6%**

Table 5: All the features include sentence embedding, context embedding, turn position embedding, sentiment and character embedding. The hybrid RCNN model with all the features performed the best on the ANNSET. Baseline models in the upper section also used all the features but didn't perform as good as the hybrid RCNN.

As shown in Table 5, the hybrid RCNN with all the features (sentence embedding, context embedding, turn position embedding, sentiment and character embedding) reached the highest accuracy (74.8%) and F1 (59.6%). Baseline models in the upper section of Table 5 also used all the features but didn't perform as good as the hybrid RCNN. We further performed ablation study on the hybrid RCNN to discover different features' impact on the model's performance. We experimented with four different context embedding methods, 1) CNN, 2) the mean of word embeddings, 3) RNN (the output of the RNN was

the RCNN's initial hidden state), and 4) tf-idf. We found RNN achieved best result (74.4%) and F1 (59.3%). The experimental results suggest incorporating context improved the model performance slightly but not significantly. This may be because in persuasion conversations, sentences are relatively long and contain complex semantic meanings, which makes it hard to encode the context information. This suggests we develop better methods to extract important semantic meanings from the context in the future. Besides, all three sentence-level features improved the model's F1. Although the sentiment feature only has three dimensions, it still increased the model's F1 score.

To further analyze the results, we plotted the confusion matrix for the best model in Fig. 5 in Appendix. We found the main error comes from the misclassification of *Personal story*. Sometimes sentences of *Personal story* were misclassified as *Emotion appeal*, because a subjective story can contain sentimental words, which may confuse the model. Besides, *Task-related inquiry* was hard to classify due to the diversity of inquiries. In addition, *Foot-in-the-door* strategy can be mistaken for *Logical appeal*, because when using *Foot-in-the-door*, people would sometimes make logical arguments about the small donation, such as describing the tangible effects of the small donation. For example, the sentence "Even five cents can help save children's life." also mentioned the benefits from the small donation. Besides, certain sentences of *Logical appeal* may contain emotional words, which led to the confusion between *Logical appeal* and *Emotion appeal*. In summary, due to the complex nature of human-human typing dialogues, one sentence may convey multiple meanings, which led to misclassifications.

7 Donation Outcome Analysis

After identifying and categorizing the persuasion strategies, the next step is to analyze the factors that contribute to the final donation decision. Specifically, understanding the effects of the persuader's strategies, the persuadee's personal backgrounds, and their interactions on donation can greatly enhance the conversational agent's capability to engage in personalized persuasion. Given the skewed distribution of intended donation amount from the persuadees, the outcome variable was dichotomized to indicate whether they donated or not (1 = making any amount of

donation and 0 = none). Duplicate survey data from participants who did the task more than once were removed before the analysis, and for such duplicates, only data from the first completed task were retained. This pruning process resulted in an analytical sample of 252 unique persuadees in the ANNSET. All measured demographic variables and psychological profile variables were entered into logistic models. Results are presented in Section A.2 in Appendix. Our analysis consisted of three parts, including the effects of persuasion strategies on the donation outcome, the effects of persuadees' psychological backgrounds on the donation outcome, and the interaction effects among all strategies and personal backgrounds.

7.1 Persuasion Strategies and Donation

Overall, among the 10 persuasion strategies, ***Donation information* showed a significant positive effect on the donation outcome** ($p < 0.05$), as shown in Table 8 in Appendix. This confirms previous research which showed efficacy information increases persuasion. More specifically, because *Donation information* gives the persuadee step-by-step instructions on how to donate, which makes the donation procedure more accessible and as a result, increases the donation probability. An alternative explanation is that persuadees with a strong donation intention were more likely to ask about the donation procedure, and therefore *Donation information* appeared in most of the successful dialogues resulting in a donation. These compounding factors led us to further analyze the effects of psychological backgrounds on the donation outcome.

7.2 Psychological Backgrounds and Donation

We collected data on demographics and four types of psychological characteristics, including moral foundation, decision style, Big-Five personality, and Schwartz Portrait Value, to analyze what types of people are more likely to donate and respond differently to different persuasive strategies.

Results of the analysis on demographic characteristics in Table 11 show that **the donation probability increases as the participant's age increases** ($p < 0.05$). This may be due to the fact that older participants may have more money and may have children themselves, and therefore are more willing to contribute to the children's charity. The Big-Five personality analysis shows that **more agreeable participants are more likely to**

donate ($p < 0.001$); the moral foundation analysis shows that **participants who care for others more have a higher probability for donation** ($p < 0.001$); the portrait value analysis shows that **participants who endorse benevolence more are also more likely to donate** ($p < 0.05$). These results suggest people who are more agreeable, caring about others, and endorsing benevolence are in general more likely to comply with the persuasive request (Hoover et al., 2018; Graham et al., 2013). On the decision style side, **participants who are rational decision makers are more likely to donate** ($p < 0.05$), whereas intuitive decision makers are less likely to donate.

Another observation reveals participants' inconsistent donation behaviors. We found that some participants promised to donate during the conversation but reduced the donation amount or didn't donate at all in the end. In order to analyze these inconsistent behaviors, we selected the 236 persudees who agreed to donate in the ANNSET. Among these persuadees, 11% (22) individuals reduced the actual donation amount and 43% (88) individuals did not donate. Also, there are 3% (7) individuals donated more than they mentioned in the conversation. We fitted the Big-Five traits score and the inconsistent behavior with a logistic regression model. The results in Table 9 in Appendix show that people who are more agreeable are more likely to match their words with their donation behaviors. But since the dataset is relatively small, the result is not significant and we should caution against overinterpreting these effects until we obtain more annotated data.

7.3 Interaction Effects of Persuasion Strategies and Psychological Backgrounds

To provide the necessary training data to build a personalized persuasion agent, we are interested in assessing not only the main effects of persuasion strategies employed by human persuaders, but more importantly, the presence of (or lack of) heterogeneity of such main effects on different individuals. In the case where the heterogeneous effects were absent, the task of building the persuasive agent would be simplified because it wouldn't need to pay any attention to the targeted audience's attribute. Given the evidence shown in personalized persuasion, our expectation was to observe variations in the effects of persuasion strategies

conditioned upon the persuadee's personal traits, especially the four psychological profile variables identified in the previous analysis (i.e., agreeableness, endorsement of care and benevolence, and rational decision making style).

Table 12, 13 and 10 present evidence for heterogeneity, conditioned upon the Big-Five personality traits, the moral foundation scores and the decision style. For example, although *Source-related inquiry* does not show a significant main effect averaged across all participants, it showed a significant positive effect on the donation probability of participants who are more open ($p < 0.05$). This suggests **when encountering more open persuadees, the agent can initiate *Source-related inquiry* more**.

Besides, ***Personal-related inquiry* significantly increases the donation probability of people who endorse freedom and care** ($p < 0.05$), **but is negatively associated with the donation probability of people who endorse fairness and authority**. Given the relatively small dataset, we caution against overinterpreting these interaction effects until further confirmed after all the conversations in our dataset were content coded. With that said, the current set of evidence supports the presence of heterogeneity in the effects of persuasion strategies, which provide the basis for our next step to design a personalized persuasive system that aims to automatically identify and tailor persuasive messages to different individuals.

8 Ethical Considerations

Persuasion is a double-edged sword and has been used for good or evil throughout the history. Given the fast development of automated dialogue systems, an ethical design principle must be in place throughout all stages of the development and evaluation. As the Roman rhetorician Quintilian defined a persuader as "a good man speaking well", when developing persuasive agents, building an ethical and good intention that benefits the persuadees must come before designing and engineering the conversational capability to persuade. For instance, we choose to use the donation task as a first step to develop a persuasive dialogue system because the relatively simple task involves persuasion to benefit children. Other persuasive contexts can consider designing persuasive agents to help individuals fulfill their goals such as engaging in more exercises or sustaining environmen-

tally friendly actions. Second, when deploying the persuasive agents in real conversations, it is important to keep the persuadees informed of the nature of the dialogue system so they are not deceived. By revealing the identity of the persuasive agent, the persuadees need to have options to communicate directly with the human team behind the system. Similarly, the purpose of the collection of persuadees personal information and analysis on their psychological traits must be clearly communicated to the persuadees and the use of their data requires active consent procedure. Lastly, the design needs to ensure that the generated responses are appropriate and nondiscriminative. This requires continuous monitoring of the conversations to make sure the conversations comply with both universal and local ethical standards.

9 Conclusions and Future Work

A key challenge in persuasion study is the lack of high-quality data and the interdisciplinary research between computational linguistics and social science. We proposed a novel persuasion task, and collected a rich human-human persuasion dialogue dataset with comprehensive user psychological study and persuasion strategy annotation. We have also shown that a classifier with three types of features (sentence embedding, context embedding and sentence-level features) can reach good results on persuasion strategy prediction. However, much future work is still needed to further improve the performance of the classifier, such as including more annotations and more dialogue context into the classification. Moreover, we found evidence about the interaction effects between psychological backgrounds and persuasion strategies. For example, when facing participants who are more open, we can consider using the *Source-related inquiry* strategy. This project lays the groundwork for the next step, which is to design a user-adaptive persuasive dialogue system that can effectively choose appropriate strategies based on user profile information to increase the persuasiveness of the conversational agent.

Acknowledgments

References

Elisabeth André, Thomas Rist, Susanne Van Mulken, Martin Klesen, and Stefan Baldes. 2000. The automated design of believable dialogues for animated presentation teams. *Embodied conversational agents*, pages 220–255.

Timothy W Bickmore, Dina Utami, Robin Matsuyama, and Michael K Paasche-Orlow. 2016. Improving access to online health information with conversational agents: a randomized controlled experiment. *Journal of medical Internet research*, 18(1).

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Chandrakant Bothe, Sven Magg, Cornelius Weber, and Stefan Wermter. 2018. Conversational analysis using utterance-level attention-based bidirectional recurrent neural networks. *Proc. Interspeech 2018*, pages 996–1000.

J. Cieciuch and E. Davidov. 2012. A comparison of the invariance properties of the pvq-40 and the pvq-21 to measure human values across german and polish samples. *Survey Research Methods*, 6(1):37–48.

Arie Dijkstra. 2008. The psychology of tailoring-ingredients in computer-tailored persuasion. *Social and personality psychology compass*, 2(2):765–784.

Brian J Fogg. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5.

CJ Hutto Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*.

Lewis R. Goldberg. 1992. The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1):26–42.

Thomas F Gordon. 1993. The pleadings game. *Artificial Intelligence and Law*, 2(4):239–292.

Arthur C Graesser, Haiying Li, and Carol Forsyth. 2014. Learning by communicating in natural language with conversational agents. *Current Directions in Psychological Science*, 23(5):374–380.

J. Graham, B. A. Nosek, J. Haidt, R. Iyer, S. Koleva, and P. H. Ditto. 2011. Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2):366–385.

Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P Wojcik, and Peter H Ditto. 2013. Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology*, volume 47, pages 55–130. Elsevier.

Shih S. I. Hamilton, K. and S. Mohammed. 2016. The development and validation of the rational and intuitive decision styles scale. *Journal of personality assessment*, 98(5):523–535.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343.

Sally Hibbert, Andrew Smith, Andrea Davies, and Fiona Ireland. 2007. Guilt appeals: Persuasion knowledge and charitable giving. *Psychology & Marketing*, 24(8):723–742.

Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*, pages 11–21.

Christopher Thomas Hidey and Kathleen McKeown. 2018. Persuasive influence detection: The role of argument sequencing. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Joe Hoover, Kate Johnson, Reihane Boghrati, Jesse Graham, and Morteza Dehghani. 2018. Moral framing and charitable donation: Integrating exploratory social media analyses and confirmatory experimentation. *Collabra: Psychology*, 4(1).

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Matthew W Kreuter, Victor J Strecher, and Bernard Glassman. 1999. One size does not fit all: the case for tailoring print materials. *Annals of behavioral medicine*, 21(4):276.

Klaus Krippendorff. 2004. Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research*, 30(3):411–433.

Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for text classification. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 2267–2273. AAAI Press.

Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? end-to-end learning of negotiation dialogues. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2443–2453.

Ewa Luger and Abigail Sellen. 2016. Like having a really bad pa: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5286–5297. ACM.

Stephanie Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument strength is in the eye of the beholder: Audience effects in persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, volume 1, pages 742–753.

Alexander Miller, Will Feng, Dhruv Batra, Antoine Bordes, Adam Fisch, Jiasen Lu, Devi Parikh, and Jason Weston. 2017. Parlai: A dialog research software platform. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 79–84.

Rita Orji and Karyn Moffatt. 2018. Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health informatics journal*, 24(1):66–91.

Richard E Petty and John T Cacioppo. 1986. The elaboration likelihood model of persuasion. In *Communication and persuasion*, pages 1–24. Springer.

Alec Radford, Rafal Jozefowicz, and Ilya Sutskever. 2017. Learning to generate reviews and discovering sentiment. *arXiv preprint arXiv:1704.01444*.

Barbara K Rimer and Matthew W Kreuter. 2006. Advancing tailored health communication: A persuasion and message effects perspective. *Journal of communication*, 56:S184–S201.

Carol A Scott. 1977. Modifying socially-conscious behavior: The foot-in-the-door technique. *Journal of Consumer Research*, 4(3):156–164.

Weiyan Shi and Zhou Yu. 2018. Sentiment adaptive end-to-end dialog systems. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1509–1519.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web*, pages 613–624. International World Wide Web Conferences Steering Committee.

Diyi Yang, Jiaao Chen, Zichao Yang, Dan Jurafsky, and Eduard Hovy. 2019. Lets make your request more persuasive: Modeling persuasive strategies via semi-supervised neural nets on crowdfunding platforms. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3620–3630.

Zhou Yu, Xinrui He, Alan W Black, and Alexander I Rudnicky. 2016a. User engagement study with virtual agents under different cultural contexts. In *International Conference on Intelligent Virtual Agents*, pages 364–368. Springer.

Zhou Yu, Ziyu Xu, Alan W Black, and Alexander Rudnicky. 2016b. Strategy and policy learning for non-task-oriented conversational systems. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 404–412.

Tangming Yuan, David Moore, and Alec Grierson. 2008. A human-computer dialogue system for educational debate: A computational dialectics approach. *International Journal of Artificial Intelligence in Education*, 18(1):3–26.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2204–2213.

A Appendices

A.1 Annotation Scheme

Table 6 and 7 show the annotation schemes for selected persuadee acts and persuader acts respectively. For the full annotation scheme, please refer to https://gitlab.com/ucdavisnlp/persuasionforgood. In the persuader's annotation scheme, there is a series of acts related to **persuasive proposition** (*proposition of donation*, *proposition of amount*, *proposition of confirmation*, and *proposition of more donation*). In general, **proposition** is needed in persuasive requests because the persuader needs to clarify the suggested behavior changes. In our specific task, donation propositions have to happen in every conversation regardless of the donation outcome, and therefore is not influential on the final outcome. Further, its high frequency might dilute the results. Given these reasons, we didn't consider propositions as a strategy in our specific context.

Category	Description
Ask org info	Ask questions about the charity
Ask donation procedure	Ask questions about how to donate
Positive reaction	Express opinions/thoughts that may lead to a donation
Neutral reaction	Express opinions/thoughts neutral towards a donation
Negative reaction	Express opinions/thoughts against a donation
Agree donation	Agree to donate
Disagree donation	Decline to donate
Positive to inquiry	Show positive responses to persuader's inquiry
Negative to inquiry	Show negative responses to persuader's inquiry

Table 6: Descriptions of selected important **persuadee** dialogue acts.

A.2 Donation Outcome Analysis Results

We used ANNSET for the analysis except for Fig. 4 and Table 11. Estimated coefficients of the logistic regression models predicting the donation probability (1 = donation, 0 = no donation) with different variables are shown in Table 8, 9, 10, 11,

Category	Description
Proposition of donation	Propose donation
Proposition of amount	Ask the specific donation amount
Proposition of confirmation	Confirm donation
Proposition of more donation	Ask the persuadee to donate more
Experience affirmation	Comment on the persuadee's statements
Greeting	Greet the persuadee
Thank	Thank the persuadee

Table 7: Descriptions of selected important non-strategy **persuader** dialogue acts.

12, and 13. Two-tailed tests are applied for statistical significance where $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$.

Persuasion Strategy	Coefficient
Logical appeal	0.06
Emotion appeal	0.03
Credibility appeal	-0.11
Foot-in-the-door	0.06
Self-modeling	-0.02
Personal story	0.36
Donation information	0.31*
Source-related inquiry	0.11
Task-related inquiry	-0.004
Personal-related inquiry	0.02

Table 8: **Associations between the persuasion strategies and the donation (dichotomized)**. $*p < 0.05$. ANNSET was used for the analysis.

Big-Five	Coefficient
extrovert	0.22
agreeable	-0.34
conscientious	-0.27
neurotic	-0.11
open	-0.19

Table 9: **Associations between the Big-Five traits and the inconsistent donation behavior** (dichotomized, 1 = inconsistent donation behavior, 0 = consistent behavior). $*p < 0.05$. ANNSET was used for the analysis.

A.3 Classification Confusion Matrix
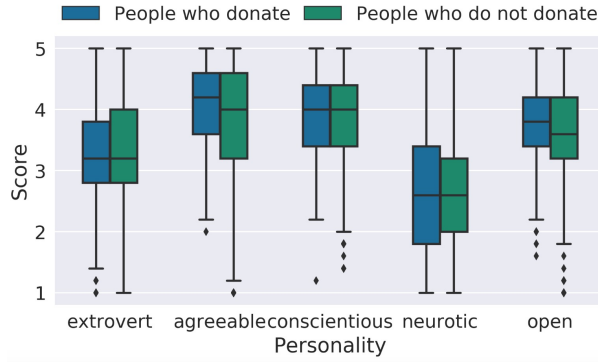
Fig. 5 shows the classification confusion matrix.

Figure 4: **Big-Five traits score distribution for people who donated and didn't donate.** For all the 471 persuadees who did not donate in the PERSUASION-FORGOOD, we compared their personalities score with the other 546 persuadees who donated. The result shows that people who donated have a higher score on agreeableness and openness in the Big-Five analysis. Because strategy annotation was not involved in the psychological analysis, we used the whole dataset (1017 dialogues) for this analysis.

Decision Style by *Strategy*	Coefficient
Rational by	
Logical appeal	0.01
Emotion appeal	0.08
Credibility appeal	-0.01
Foot-in-the-door	-0.25
Self-modeling	0.007
Personal story	0.26
Donation information	0.09
Source-related inquiry	0.33
Task-related inquiry	-0.03
Personal-related inquiry	-0.03
Intuitive by	
Logical appeal	0.04
Emotion appeal	-0.07
Credibility appeal	-0.02
Foot-in-the-door	0.37
Self-modeling	0.01
Personal story	-0.27
Donation information	-0.02
Source-related inquiry	-0.43
Task-related inquiry	0.05
Personal-related inquiry	0.04

Table 10: **Interaction effects between decision style and the donation (dichotomized).** $*p < 0.05$. Coefficients of the logistic regression predicting the donation probability (1 = donation, 0 = no donation) are shown here. ANNSET was used for the analysis.

Predictor	Coefficient
Demographics	
Age	0.02*
Sex: Male vs. Female	-0.11
Sex: Other vs. Female	-0.14
Race: White vs. Other	0.28
Less Than Four-Year College vs. Four-Year College	0.16
Postgraduate vs. Four-Year College	-0.20
Marital: Unmarried vs. Married	-0.21
Employment: Other vs. Employed	0.17
Income (continuous)	-0.01
Religion: Catholic vs. Atheist	0.34
Religion: Other Religion vs. Atheist	0.21
Religion: Protestant vs. Atheist	0.15
Ideology: Liberal vs. Conservative	0.11
Ideology: Moderate vs. Conservative	-0.04
Big-Five Personality Traits	
Extrovert	-0.17
Agreeable	0.58***
Conscientious	-0.15
Neurotic	0.09
Open	-0.01
Moral Foundation	
Care/Harm	0.38***
Fairness/Cheating	0.08
Loyalty/Betrayal	0.09
Authority/Subversion	0.04
Purity/Degradation	-0.02
Freedom/Suppression	-0.13
Schwartz Portrait Value	
Conform	-0.07
Tradition	0.06
Benevolence	0.18*
Universalism	0.05
Self-Direction	-0.06
Stimulation	-0.08
Hedonism	-0.10
Achievement	-0.03
Power	-0.05
Security	0.09
Decision-Making Style	
Rational	0.25*
Intuitive	-0.02

Table 11: **Associations between the psychological profile and the donation (dichotomized).** $*p < 0.05$, $***p < 0.001$. Estimated coefficients from a logistic regression predicting the donation probability ((1 = donation, 0 = no donation)) are shown here. Because strategy annotation is not involved in the demographical and psychological analysis, we used the whole dataset (1017 dialogues) for this analysis.

A.4 Data Collection Interface
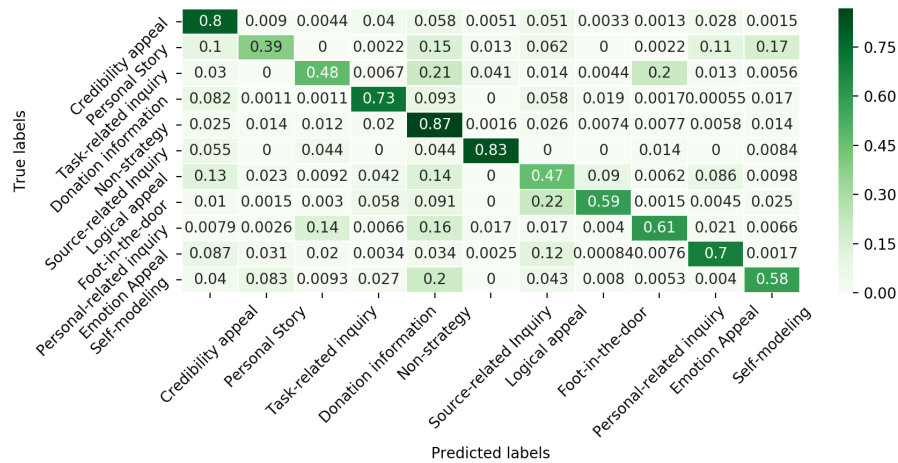
Fig. 6 and 7 shows the data collection interface.

Figure 5: Confusion matrix for the ten persuasion strategies and the non-strategy category on the ANNSET using the hybrid RCNN model with all the features.
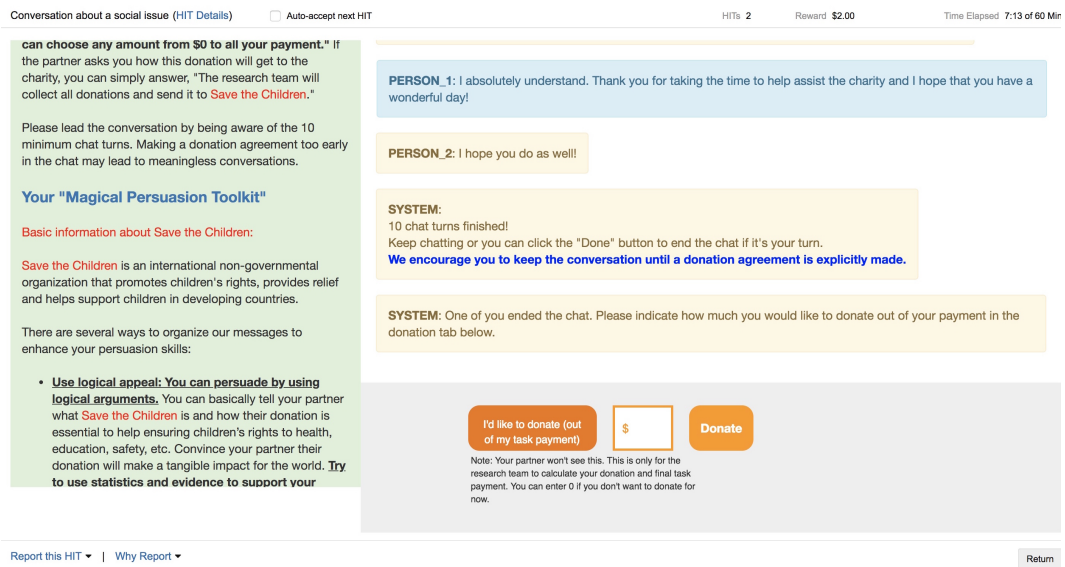

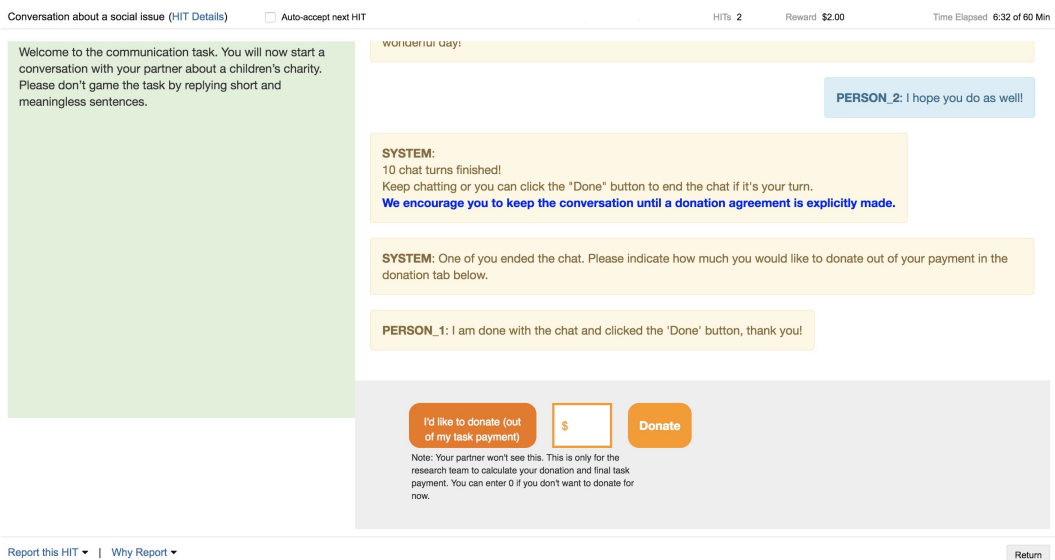Figure 6: Screenshot of the persuader's chat interface

Figure 7: Screenshot of the persuadee's chat interface

5648

Big-Five by *Strategy*	Coefficient
Extrovert by	
Logical appeal	-0.06
Emotion appeal	0.15
Credibility appeal	0.07
Foot-in-the-door	0.21
Self-modeling	-0.28
Personal story	-0.18
Donation information	-0.11
Source-related inquiry	-0.02
Task-related inquiry	-0.26
Personal-related inquiry	0.09
Agreeable by	
Logical appeal	-0.11
Emotion appeal	0.25
Credibility appeal	0.25
Foot-in-the-door	-0.02
Self-modeling	-0.30
Personal story	0.77
Donation information	0.08
Source-related inquiry	-0.84
Task-related inquiry	-0.61
Personal-related inquiry	-0.07
Neurotic by	
Logical appeal	0.12
Emotion appeal	-0.14
Credibility appeal	-0.03
Foot-in-the-door	0.05
Self-modeling	-0.20
Personal story	-0.22
Donation information	0.15
Source-related inquiry	-0.22
Task-related inquiry	0.03
Personal-related inquiry	0.23
Open by	
Logical appeal	0.13
Emotion appeal	0.21
Credibility appeal	-0.20
Foot-in-the-door	-0.97
Self-modeling	0.38
Personal story	-0.17
Donation information	-0.33
Source-related inquiry	1.21*
Task-related inquiry	0.63
Personal-related inquiry	-0.21
Conscientious by	
Logical appeal	-0.02
Emotion appeal	-0.40
Credibility appeal	-0.14
Foot-in-the-door	0.67
Self-modeling	0.34
Personal story	-0.28
Donation information	0.33
Source-related inquiry	-0.03
Task-related inquiry	0.21
Personal-related inquiry	0.06

Table 12: **Interaction effects between Big-Five personality scores and the donation (dichotomized).** $*p < 0.05$, $**p < 0.01$. Coefficients of the logistic regression predicting the donation probability (1 = donation, 0 = no donation) are shown here. ANNSET was used for the analysis.

Moral Foundation by *Strategy*	Coefficient
Care by	
Logical appeal	0.05
Emotion appeal	-0.19
Credibility appeal	0.21
Foot-in-the-door	0.03
Self-modeling	0.54
Personal story	0.12
Donation information	-0.21
Source-related inquiry	0.14
Task-related inquiry	0.09
Personal-related inquiry	1.10*
Fairness by	
Logical appeal	0.12
Emotion appeal	0.06
Credibility appeal	-0.10
Foot-in-the-door	-0.40
Self-modeling	-0.09
Personal story	-0.30
Donation information	0.06
Source-related inquiry	0.46
Task-related inquiry	0.41
Personal-related inquiry	-1.15*
Loyalty by	
Logical appeal	-0.10
Emotion appeal	-0.13
Credibility appeal	0.07
Foot-in-the-door	0.45
Self-modeling	0.04
Personal story	-0.31
Donation information	-0.25
Source-related inquiry	0.57
Task-related inquiry	-0.26
Personal-related inquiry	-0.04
Authority by	
Logical appeal	0.31
Emotion appeal	-0.12
Credibility appeal	0.10
Foot-in-the-door	-0.31
Self-modeling	0.08
Personal story	-0.19
Donation information	0.03
Source-related inquiry	-0.23
Task-related inquiry	-0.14
Personal-related inquiry	-0.86*
Purity by	
Logical appeal	-0.30
Emotion appeal	0.25
Credibility appeal	-0.15
Foot-in-the-door	-0.004
Self-modeling	-0.21
Personal story	0.43
Donation information	0.30
Source-related inquiry	-0.41
Task-related inquiry	0.31
Personal-related inquiry	0.44
Freedom by	
Logical appeal	0.10
Emotion appeal	-0.05
Credibility appeal	-0.16
Foot-in-the-door	-0.50
Self-modeling	-0.35
Personal story	0.32
Donation information	0.17
Source-related inquiry	-0.13
Task-related inquiry	-0.29
Personal-related inquiry	0.60*

Table 13: **Interaction effects between moral foundation and the donation (dichotomized).** $*p < 0.05$.