Persuasion in online communication: automation and counteraction

Samira Shaikh Shaikh
University at Albany, State University of New York, samirashaikh@gmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Artificial Intelligence and Robotics Commons, and the Social Psychology Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
Persuasion in Online Communication
– Automation and Counteraction

By

Samira Shaikh

A Dissertation
Submitted to the University at Albany, State University of New York
In Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

Department of Computer Science
2016
Persuasion in Online Communication

– Automation and Counteraction

By

Samira Shaikh

COPYRIGHT 2016
In this thesis, we studied persuasion in online communication and how to automate persuasive behavior in an autonomous chat agent. We implemented known persuasive strategies into the agent, which are based upon the strength and evaluation of the beliefs expressed by participants in conversation, to induce belief change. The foundation of our persuasive strategies comes from the summative model of attitude, where belief change leads to attitude change, and, ultimately, behavior change. Upon placing an agent in the midst of conversations, it is able to discern beliefs that are expressed by the participants in the group, and use them to ascertain participant’s opinions on topics of discussion. Using this information and drawing upon theories of influence and persuasion from social psychology, cognitive science and communication, the agent aligns participants towards or against a particular issue.

We organized the work in three phases. First, we conducted a belief elicitation study to obtain salient beliefs on a variety of social issues and used these salient beliefs to create survey instruments. Next, we programmed behaviors and strategies in the agent that were aimed at persuading individuals through online conversation as well as counteracting
persuasion by the participants. The behaviors programmed in the agents are triggered, in part, by a variety of linguistic cues emerging from the conversation, such as dialogue acts, topic, polarity and communication acts. The annotated context of conversation is used to inform the agent’s models by updating the underlying beliefs of participants in real time. Third, we ran controlled experiments with human participants to validate the chat agent in a variety of settings, including Wizard-of-Oz and autonomous agent conditions to determine the efficacy of its programmed strategies. In the validation experiments, we used pre-discussion and post-discussion surveys to determine changes in participants’ attitudes prior to and after a discussion. We showed that the agent achieved statistically significant changes in the participant’s attitudes, thus demonstrating its effectiveness in being persuasive. Through our work, we have shown that specific persuasion strategies can be automated as well as counteracted using sophisticated communication models built upon sociolinguistic and psychological theories of social influence.
ACKNOWLEDGMENTS

This dissertation is an undertaking that has spanned several formative years of my life. Over these years, I have received support, advice, encouragement and inspiration from a number of people. Foremost, I thank my advisor, Professor Tomek Strzalkowski. I am forever indebted to him for things I cannot even count. This dissertation would not have come to be if it were not for him. I hope I can someday achieve all that he has inspired me to achieve.

I am thankful to Professor Laurie Feldman. Her mentorship, encouragement and insights have given me clarity, confidence and strength. I look up to her and will continue to seek inspiration from her as a mentor and advisor.

I thank Professor Neil Murray, for his advice and support. My work has benefited greatly by working and collaborating with several excellent people during these years, including Ting Liu, Veena Ravishankar, Jennifer Crowley, Ken Stahl, Nick Webb, Sharon Small, Mike Ferguson and Ching-Sheng Lin. These pages are few and the list of people to thank is potentially endless. I hope I get the opportunity to thank everyone I want to in person.

I do want to thank my family. My parents made me who I am. I am extremely grateful to them for everything they have done for me. Finally, I thank my husband, Ashish. His support and understanding are everything I could ask for in a partner. He motivates me to achieve much more than I ever hoped for.

I dedicate this thesis to my father and mother.
# TABLE OF CONTENTS

## Chapter 1: Introduction and Research Questions

1.1 Introduction .................................................. 1  
1.2 Research Questions ......................................... 2  
1.3 Motivation .................................................... 6  
1.4 Terminology .................................................. 8  
1.5 Ethical Considerations ....................................... 9  
1.6 Structure of this thesis ..................................... 11

## Chapter 2: Literature Review

2.1 Introduction .................................................. 12  
2.2 Persuasion and Models of Attitude Change ............... 13  
  2.2.1 Functional Approaches towards Attitude Change ... 13  
  2.2.2 Elaboration Likelihood Models towards Attitude Change ... 15  
  2.2.3 Summative Model of Attitude Change ................ 16  
2.3 Artificial Agents ............................................ 21  
2.4 Computational Analysis of Discourse ..................... 28  
  2.4.1 Dialogue Act Tagging .................................. 29  
  2.4.2 Modeling Sociolinguistic Phenomena in Dialogue ... 32  
2.5 Topical Positioning .......................................... 37  
2.6 Summary of this chapter .................................. 40

## Chapter 3: Belief Elicitation Study

3.1 Introduction .................................................. 42  
3.2 Motivation .................................................... 44  
3.3 Participants .................................................. 46  
3.4 Topics ........................................................ 47  
3.5 Survey Questions .......................................... 47  
3.6 Procedure ...................................................... 49  
3.7 Analysis of Data - Annotation Procedure ................ 51  
3.8 Analysis of Data - Descriptive Statistics ................. 54  
  3.8.1 Topic: Genetically modified foods .................... 55  
  3.8.2 Topic: Living together before marriage .............. 57  
  3.8.3 Topic: Being able to identify people online ......... 58  
  3.8.4 Topic: Easy access to morning after pill to people under 17 ... 60  
  3.8.5 Topic: Lowering minimum legal drinking age from 21 to 18 ... 61  
  3.8.6 Topic: Stricter airport screening procedures ......... 63  
  3.8.7 Contrastive analysis of questions across topics ....... 64  
3.9 Themes selected for inclusion in surveys ................. 64  
3.10 Themes and arguments programmed into the agent ....... 67  
3.11 Summary of this chapter .................................. 68
# Chapter 4: Method

4.1 Introduction  70  
4.2 Agent Architecture  70  
4.3 Chat Analyzer  72  
4.3.1 Communication Link Tagger  72  
4.3.2 Dialogue Act Tagger  76  
4.3.3 Belief Extractor  78  
4.3.4 Chat Context Builder  83  
4.4 Belief Modeler  86  
4.4.1 Creating Belief Models for the Agent  87  
4.4.2 Creating Belief Models for the participants  91  
4.4.3 Updating Belief Models  95  
4.5 Behavior Selector  96  
4.5.1 Make Conventional Opening Statement  97  
4.5.2 Make Conventional Closing Statement  98  
4.5.3 Make Other Conventional Statement  98  
4.5.4 Make Assertion Opinion Statement  100  
4.5.5 Acknowledge Agree Accept  103  
4.5.6 Agree with Reinforcement  103  
4.5.7 Disagree with Reasons  104  
4.6 Proactive Behaviors  106  
4.7 Counter-persuasion strategies  108  
4.8 Summary of this chapter  111  

# Chapter 5: Evaluating the Agent

5.1 Introduction  112  
5.2 Experiment Design  112  
5.2.1 Wizard of Oz and Autonomous Agent  112  
5.2.2 Small Groups and Majority-Minority Influence  114  
5.3 Participants and Schedule  119  
5.4 Procedures  121  
5.4.1 Participant laboratory experience  123  
5.4.2 Wizard training  127  
5.4.3 Measures and Metrics  130  
5.5 Data Analysis and Results – WOZ experiments  132  
5.5.1 Descriptive Statistics  132  
5.5.2 Perception Ratings  135  
5.5.3 Persuasion Effectiveness  137  
5.5.4 Wizard debriefing results  147  
5.5.5 Conclusion from WOZ experiments  148
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6 Data Analysis and Results - Autonomous agent experiments</td>
<td>149</td>
</tr>
<tr>
<td>5.6.1 Refining the algorithm</td>
<td>149</td>
</tr>
<tr>
<td>5.6.2 Descriptive Statistics</td>
<td>153</td>
</tr>
<tr>
<td>5.6.3 Perception Ratings</td>
<td>156</td>
</tr>
<tr>
<td>5.6.4 Persuasion Effectiveness</td>
<td>157</td>
</tr>
<tr>
<td>5.6.5 Qualitative interviews with participants</td>
<td>163</td>
</tr>
<tr>
<td>5.7 Summary of this chapter</td>
<td>164</td>
</tr>
<tr>
<td>Chapter 6: Conclusion</td>
<td>166</td>
</tr>
<tr>
<td>6.1 Introduction</td>
<td>166</td>
</tr>
<tr>
<td>6.2 Future Directions</td>
<td>166</td>
</tr>
<tr>
<td>6.2.1 Further Validation against larger sample size</td>
<td>166</td>
</tr>
<tr>
<td>6.2.2 Further validation using a control condition</td>
<td>167</td>
</tr>
<tr>
<td>6.2.3 Inoculation Experiments</td>
<td>167</td>
</tr>
<tr>
<td>6.2.4 Validation “in the wild”</td>
<td>168</td>
</tr>
<tr>
<td>6.3 Contributions of this thesis</td>
<td>169</td>
</tr>
<tr>
<td>References</td>
<td>170</td>
</tr>
<tr>
<td>Appendix A: Belief Elicitation survey</td>
<td>185</td>
</tr>
<tr>
<td>Appendix B: Pre-discussion survey</td>
<td>187</td>
</tr>
<tr>
<td>Appendix C: Post-discussion survey</td>
<td>207</td>
</tr>
<tr>
<td>Appendix D: Debriefing form</td>
<td>209</td>
</tr>
<tr>
<td>Appendix E: Wilcoxon signed-rank test calculations</td>
<td>211</td>
</tr>
<tr>
<td>Appendix F: Glossary and definitions of Dialogue Act Tags</td>
<td>214</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction and Research Questions

1.1 Introduction

Machine intelligence breaks new ground with each passing day, achieving feats like winning against humans in the game of Go and Jeopardy! (Markoff, 2011; Wood, 2016). However, the goal of implementing social competence in machines so that they may achieve specific social goals remains elusive. Persuasion, the ability to tailor messaging to the target audience is closely related to social competence (Grass and Seiter, pp 48:49). Persuasive technology is an interdisciplinary field of research focusing on the design, implementation and validation of technology aimed at creating intelligent machines that generate behaviors to change attitudes and behaviors of its users. Effective persuasion and social influence by intelligent machines depend not only on appropriate interventions but also on the initial beliefs of users. Changing behavior and beliefs by social interaction has widespread applications in various domains such as healthcare, education and marketing. For example, in the health domain, persuasive technology has been used to create virtual aerobics trainers for motivating workouts in overweight adults (Davis and Bobick, 1998); smoking cessation apps to cue smoking reduction programs to users (Goh et al, 2009) and nutrition coaches to provide balanced diet regimens (Purpura et al. 2011). In the education domain, persuasive technology has been used to design software to motivate young children to practice the target behavior of reading and writing (Lucero, Zuloaga and Muñoz, 2006) and in the marketing domain, the design of websites that engage and motivate users to purchase products has received wide attention (c.f. Fogg, 2003 for an exhaustive survey).
In using technology to persuade, we delegate a central function of human communication to a computer. This central function is intentional and goal-directed communication - deliberate, purposeful messaging with a clear outcome in mind. When communicating, people, both intentionally and unintentionally, reveal and promote certain beliefs and behaviors. Consequently, persuasion is inherent in social interaction. The existing persuasive technologies we have reviewed, whilst being quite effective, can only be used to deliver targeted interventions in specific contexts. Our goal in this thesis is to automate the very process of persuasive communication, by designing a system which can purposefully communicate, without any restrictions on domain or genre or task, and which has the clear intention of persuading the recipients of its messaging. In doing so, we advance the field of artificial intelligence and human-computer dialogue. We investigate how models of social phenomena - specifically persuasion strategies - may be automated in an artificial autonomous agent, in the form of two overarching research questions.

1.2 Research Questions

Our work investigates two main research questions explicated below:

1. *Can specific persuasive strategies be automated in a virtual chat agent?*

2. *Can active persuasion by individuals during conversation be detected and counteracted by such an agent?*

To expand further, we want to know how we can deploy an agent in an online multi-party chat room where its goal will be to persuade those participants who hold an opposing view than its own and empirically test its efficacy in such settings. We specifically wish to test the agent in settings of majority-minority influence. The agent holds the minority opinion on the topic of discussion, while the majority of participants will hold an opposing opinion.
The motivation for selecting minority influence role for the agent comes from social influence research (Martin, Martin, Smith and Hewstone, 2007; Gardikiotis, 2011) who have shown that minority influence leads to more consistent attitude and behavior change in the receivers of persuasive messaging. A majority-minority setting may inherently lead to situations where the participants who hold majority opinion may also try to persuade the minority. Thus, for our second research question, we want the agent to detect when other participants in the conversation are being persuasive, as this is the first step to actively counteracting such persuasion.

The nature of our research introduces a set of challenges that needs to be addressed, in addition to the main undertaking of automating persuasion. We describe below the major challenges we faced and our solution for these challenges.

**Challenge 1: The agent must remain undetected as a computer program.** If the participants in conversation ascertain that a computer agent is acting as a participant, they may not respond in a socially natural manner and the validity of our experimental data would be compromised. Shechtman & Horowitz (2003) studied the perception of human-computer interaction when users thought they were using the computer to chat with a real person and when they thought they were only interacting with a machine. They showed that when participants believed they were talking to a person, the participants exhibited many more of the kinds of behaviors associated with establishing the interpersonal nature of a relationship. On the other hand, Fogg (2003) determined that people give more credibility and trustworthiness to computers because they often view computers as incorporating expertise. In revealing the presence of a computer agent to the participants in our experiments we may potentially confound our experimental findings with such
preconceptions and biases. Moreover, the aim of this research is to model human persuasive capability in a machine, not to ascertain what it would take to create a persuasive machine.

To that end, we aim for the agent to remain undetected as a computer program for the entire length of the conversation. We note that then this research may be construed as a variant of the Turing Test (Turing, 1950), which aims to answer the fundamental question “Can machines think?” or its more relaxed equivalent, the Loebner Prize (Shieber, 1993).

One reason we chose the multiparty discourse setting for our research is that the presence of multiple participants distributes the onus of conversation across all participants, not just the agent. Our hypothesis is that the presence of multiple participants will act as a safeguard against easy detection of the agent among their midst (Dohsaka et al., 2009).

Indeed, most of the existing work on conversation agents has focused on agents that support only one user in each interactive session. However, people tend to organize themselves in groups such as teams of coworkers, family and networks of friends. We therefore deployed the agent in multi-party conversation settings and this, in part, overcame the challenge of being undetected. The experiments to validate the agent thus required the use of deception, and we obtained the necessary institutional review board permissions to do so.

**Challenge 2:** The agent must behave in a socially appropriate manner. This challenge overlaps significantly Challenge 1 mentioned above. If the agent is to remain undetected, then it must also behave in a socially capable manner, which entails that the agent should adhere to conversational norms in the dialogue. Studies have shown that agents, and indeed humans, with social capabilities can achieve significantly better task success and perception ratings compared to agents that perform the same task related behaviors.
without social skills (Kumar et al., 2011). Similarly, Dohsaka et al. (2009) found that an agent’s use of emphatic expressions improved user satisfaction and user rating of the agent, in the context of multi-party thought-evoking dialogue. The user ratings of the agent included user’s perceived closeness, likability and caring from the agent, and the user’s feeling of being supported by the agent. Accordingly, we ensured that the agent has at its disposal a set of social behaviors consistent with conversational norms to overcome this challenge.

Challenge 3: The agent must be able to output natural language expressions in real time. This challenge is also related inherently to Challenge 1. If the agent is to remain undetected, the utterances that it makes during conversation must be natural language expressions of the sort expected from human participants. Of course, the aim of the research is that the expressions generated by the agent will be persuasive messages that follow an algorithmic principle. However, our goal in this thesis is not to examine the problem of natural language generation, which may require language production from formal representations (Reiter et al., 2005). Instead, we used a simple template-based content generation model, in which the goal is to select from among preprogrammed templates, which are augmented with natural language expressions generated by humans. The human-generated natural language expressions were acquired during a belief elicitation study, which was conducted, in part, to create pre-discussion and post-discussion surveys for our validation experiments. During the belief elicitation study, subjects were asked to respond to questions related to socially relevant topics in free-form text. Beyond the experimental settings of this research, such natural-language expressions could be harvested from the Internet and social media.
Our goal is not, therefore, to create an artificial agent capable of passing the Turing test or the Loebner prize. Rather, our goal is to define specific human persuasive strategies that can be programmed into an agent who can then persuade participants to its own view during the course of conversation. Our motivation to do so is described in the next section.

1.3 Motivation

Artificial agents become increasingly a part of everyday life, such as agents for preventive medicine, education and targeted advertising (Davis and Bobick, 1998; Lucero, Zuloaga and Muñoz, 2006; Fogg, 2003). A lot of online communication occurs today between groups of individuals (big and small). The use of social media and online interactions in starting and sustaining political movements of recent times, such as the Arab spring (Marzouki et al., 2012; Harlow, 2013; Al-Rawi, 2014) cannot be underestimated. Other applications where multi-party interaction may take place in a computer-mediated communication (CMC) setting are collaborative learning (Kumar et al., 2007) and collaborative work (Isbister et. al., 2000; Harris and Rudnicky, 2007) as well as multi-player online games (Tse et al., 2007). In collaborative work, group discourse and group decision-making takes place in CMC settings. Research has shown that groups communicating via computers produce more polarized decisions (Lea and Spears, 1991) and exhibit more uninhibited behavior (Kiesler et al. 1983). Additionally, the anonymity and disinhibition afforded by CMC leads to a free exchange of ideas and encourages expression of minority viewpoints (Haines et al., 2014). In collaborative decision-making, participants exchange information and offer different points of view in order to achieve a common goal: the resolution of a problem. Group participants discuss a situation at hand, explain issues, contribute different opinions and present various propositions that may contradict other participants’ views. Hence,
certain participants have to be persuaded to change some of their attitudes (beliefs, goals, actions) when they face contributions that contradict their previous beliefs or their understanding of the situation. If a person judges that the arguments presented are stronger than the arguments supporting her previous contention, there are chances that she will adopt the new attitude. Persuasion then entails making a participant realize the strength of certain arguments over her own previous opinion and inciting action on the basis of such attitude change. As such, persuasive communication plays a large role in group-decision making and collaborative work, and in the context of CMC, it is feasible that interlocutors may indeed be virtual agents instead of humans. Consequently, applications of technology that can actively engage in autonomous persuasion strategies in multi-party CMC settings and counteract active persuasion by others are far-reaching and broad.

Many theories in sociology, philosophy, psychology and communication have been formulated to improve persuasive communication. Sociologists, for instance, concentrate on the identification of features of communication that can influence persuasiveness, studying the receiver's motivation, intentions and values and the correlation with the output of persuasion (e.g. Taillard and Giscoppa, 2013). Philosophers have long studied rhetoric to formalize the techniques of selecting and presenting data in a persuasive way (c.f. Kennedy, 2015 for an exhaustive survey). The underpinnings of our work, however, derive from the psychology of social influence and theories of attitude change from the communication literature, pioneered by Cialdini (1987) and Azjen (2012).

The aim of this research is to develop and test novel techniques to improve human-computer interaction by improving its persuasiveness. Indeed, the research presented here concentrates on the development of persuasion through human-computer dialogue. With
the mass-adoption of computer-mediated communication technologies for group interaction, there is an unprecedented opportunity and need for intelligent agents to support interactive situations involving multiple participants. For example, Huang and Lin (2011) developed a persuasive sales clerk, which adapts to different characteristics of online buyers in e-stores. Al Mahmud et al. (2011) created an intelligent agent that helps persuade family members to conserve energy in their home. Cassell and Bickmore (2003) built a virtual embodied agent to establish collaborative trusting relationships with human users, with applications designed for financial transactions and advice-giving. The human-computer interfaces of the future will be geared towards inducing the user to perform certain actions in the real world (Guerini, Stock and Zancanaro, 2003). This unprecedented opportunity and need for such persuasive intelligent technology is the motivation for this thesis.

1.4 Terminology

At the outset, it is important to define what we mean by persuasion. Persuasion has been extensively studied in literature by researchers (Cialdini, 1987, Perloff, 2010; O’Keefe, 2015), who make explicit distinctions between persuasion and coercion or manipulation. Persuasion requires the assumption of free will on the part of the receiver as opposed to coercion or manipulation (O’Keefe, 2015).

Coercion involves the use of force, where the free will of the receiver is compromised. Philosophers define coercion as a technique for forcing people to act as the coercer wants them to act, and presumably contrary to their preferences. It usually employs a threat of some dire consequence if the actor does not do what the coercer demands (Feinberg, 1998, p. 387).
Manipulation differs from persuasion and coercion as it involves distorting or withholding truth so that the receiver cannot freely and reasonably make a correct choice. Persuasion also requires clear intent on part of the communicator, but the receiver is free to question the persuader or offer contrasting opinions. In this work, even though there is ostensible deception undertaken during experiments to avoid confounding the experimental results, there is no coercion or manipulation involved. In our Literature Review (Chapter 2), we expand further upon the definition of persuasion adopted in this thesis.

1.5 Ethical Considerations

There are admittedly ethical considerations of designing and deploying persuasive technology. Ethicists, who have long studied persuasion, posit that persuasion apparently distributes responsibility between the persuader and the persuaded (Anderson, 1971). Some communication theorists have declared that “persuasion is ethically neutral” (McCroskey, 1972). They argue that in most simple cases, where one person is persuading another, all involved parties share full moral and ethical accountability for the outcome.

In their work, Berdichevsky and Neuenschwander (1999) established the first set of principled guidelines for the design and implementation of persuasive technology. There is indeed potential for any technology to have unintended consequences or unpredictable outcomes and such consequences or outcomes might be negative or unethical. In Figure 1.5.1, we show the levels of ethical responsibility on the designer of persuasive technology envisaged by Berdichevsky and Neuenschwander (1999).
As we can see, the designer is not held responsible for any unintended unethical outcomes, which cannot be reasonably predicted. However, the designer is responsible and at fault for any intended unethical outcomes or unintended unethical outcomes which could have been reasonably predicted. In our proposed research, there are no unethical outcomes that are intended or reasonably predictable, because our goal is simply to engage participants in online conversation on non-controversial socially relevant topics of discussion. Any resulting attitude and behavior change on those topics would not produce unethical outcomes that we can reasonably predict beforehand.
1.6 Structure of this thesis

This thesis has six chapters, including this introduction. The second chapter presents a review of prior work and provides an overview of the state-of-the-art in the field. In Chapter 2, we also present the foundation of this thesis from social influence literature and theories of attitude change. We describe techniques from natural language processing and computational linguistics such as belief modeling, sentiment analysis, dialogue act tagging and determining the positions of speakers on salient topics of conversation that have bearing upon successful realization of the agent. In Chapter 3, we present details about a belief elicitation study, which was conducted to obtain naturally occurring statements of text from a population of subjects about a variety of topics. We elaborate on motivation for conducting such a study, as a necessary step to generate pre-discussion and post-discussion surveys and to elicit natural-language responses, which may be used by the agent during conversation. Chapter 4 describes our proposed method. We first present the architecture of the agent and provide technical and implementation details. The validation experiments conducted and analysis of data and findings are described in Chapter 5. We demonstrate the efficacy of the agent using multiple metrics to measure success. Notably, we find the results on the persuasiveness metric show statistical significance as measured against pre-test/post-test scores of participants. We conclude with a discussion of possible future directions, contributions and conclusions in Chapter 6.
Chapter 2: Literature Review

2.1 Introduction

In this chapter, we shall review relevant literature in research areas that bear upon this thesis. The sections that follow deal with four main areas of interest.

Section 2.2 pertains to the literature in communication and social science research related to persuasion. The arguments in this section will provide the motivation for choosing the summative model of attitude for our research, which provides a path to behaviors, from underlying attitudes and beliefs of the message recipients. We provide our motivation for selecting the summative model of attitude over the leading comparable approaches that have been advanced to date. Section 2.3 provides an overview of current Conversational Agent capabilities in order to assess the state of the art and how our approach differs from existing approaches. In Section 2.4, we provide details of prior research on computational approaches for modeling phenomena in discourse. In Section 2.5, we present an existing approach towards determining the positions on participants on various topics of conversation, which will serve as the basis of creating representations of participant’s mental states. Finally, we will conclude the chapter by providing our assessment of the related research areas in Section 2.6 and showing how our proposed approach is built using insights gathered from the literature.
2.2 Persuasion and Models of Attitude Change

Miller (1980) described persuasive communication as an intentional act causing response in others. Perloff (2003) gave a comprehensive definition of persuasion, which we shall adopt in this thesis. It is presented below.

“(Persuasion is) a symbolic process in which communicators try to convince other people to change their attitudes or behavior regarding an issue through the transmission of message, in an atmosphere of free choice.”

The definition implies certain elements as part of the persuasion process, namely - communicators, receivers, messages, a medium of transmission and intent to persuade. To elaborate more on the “change in attitude or behavior” part of the definition - Rokeach (1968) initially and later, Fishbein and Ajzen (1975) (among others) explored the construct of attitude to explain responses that people have towards objects or situations, when subjected to persuasion. Persuasion can therefore be understood in the context of different approaches to attitude and attitude change. Many approaches have been advanced to understand how attitudes are formed and how they may be changed. We review three major approaches to attitude change in the following sections, Functional Approaches, Elaboration Likelihood Model and Summative Model.

2.2.1 Functional Approaches towards Attitude Change

Functional Approaches have been proposed (Katz, 1960, Maio and Olson, 2000) where analysis of attitude focuses on the functions attitudes can serve for people. The functional approach is aimed at identifying why attitudes are held and what functions they serve. Under the functional approach, attitudes may be held because they serve the function of being utilitarian, ego-defensive, value-expressive etc. When an attitude serves a utilitarian
function, it reflects rewards and punishments that the attitude object supplies to the holder of the attitude. As an example, Perloff (2003) states that athletes find it utilitarian to develop a positive attitude towards a tough coach. A positive attitude can help them get along with the tough coach (reward) and minimize their chances of earning the coach’s wrath (punishment). Other examples of the utilitarian function are advertisements that stress straightforward benefits of products (“you should drink Diet Coke because it tastes great” (Perloff, 2003)). Similarly, under the Functional Approach model, an attitude can serve an ego-defensive function, where it is assumed that the attitude protects the ego through defense mechanisms such as projection or distancing oneself from the disliked attitude object (Katz, 1960). For instance, a person’s negative attitude towards groups such as homosexuals may reflect attempts to protect the ego from the threats these outgroups are perceived to pose (Herek, 1987).

Persuasion strategies under this model are either aimed at reinforcing the function of attitude, known as the matched appeal or undermining the attitude, known as the mismatched appeal. For our purpose, the implications of using functional approaches of attitude change as a basis of programming a persuasion strategy in the agent are that we would need to devise strategies that automatically ascertain the function an attitude may serve for an individual through the language used by this individual during online conversation. Moreover, people can hold the same attitude towards an attitude object even though it might serve entirely different functions. It may, thus, be difficult to assess such functions of an attitude with reliable consistency within a group of people, and more difficult still to deploy an effective persuasion strategy through an automated agent built on a Functional Approach. We wanted to build our persuasion agent so that the strategies it
employs would work regardless of the function of the attitude. Thus, we did not base our
persuasion strategies on the Functional Models of Attitude.

2.2.2 Elaboration Likelihood Models (ELM) towards Attitude Change

Elaboration Likelihood Models (Petty and Cacioppo, 1986) are based on a dual-processing
approach, which suggests that under different conditions, recipients of persuasive
messages will vary in the degree of engaged thinking about the persuasive issue referred to
as elaboration. When elaboration is extensive in a receiver, the persuasive message
contents become important. When elaboration is low, peripheral cues or heuristics – such as
whether receivers think that the communicator is likeable or credible - become important.
The more individuals are likely to think about the persuasive message, the more
elaboration arises, leading to increased potential for carefully designed messaging to be
persuasive. The key factors that determine processing of a message then, are motivation
and ability. When people are motivated or are cognitively able to seriously consider the
message, then they will process it centrally leading to higher elaboration. Otherwise, when
people lack the motivation or ability, they may process it superficially and opt for heuristics
or biased thinking (Tversky and Kahneman, 1974; Kahneman, 2003).

While proponents of the ELM argue that the model can be used to explain all possible
outcomes (Petty et al., 1993), critics have used precisely this fact to level disapproval. It is
argued in Allen and Preiss (1997, pp 117-118) that using the ELM “A persuader can
conduct a post-mortem to find out what happened but cannot forecast the impact of a
particular message”. In this research, the use of ELM for building an agent would imply that
a communicator responsible for persuasive messaging would need to estimate the degree
of thinking or elaboration likely to occur in the receivers’ minds. Since the only mode of
communication available to us is computer-mediated online conversation, if the agent is built upon the ELM model, it would need to assess the motivation and ability of individuals to process its messages through their language use. Additionally, if the receiver’s motivation is low or cognitive ability is not sufficient, then the persuasion strategy could only depend on the communicator’s peripheral cues, such as their likability and credibility (Petty et al., 1993). In such situations, the input messaging becomes secondary to surface characteristics of the communicator and the receiver’s motivation and ability.

Our ultimate goal is not to create a likable or credible agent, but rather a persuasive agent. Therefore, we do not use this model in our research. A persuasive agent may need likability and credibility characteristics in order to be successful and effective, however, these attributes are not adequate to inform programming of the underlying attitude change model.

2.2.3 Summative Model of Attitude Change

The Summative Model of Attitude (Ajzen and Fishbein, 1973; O’Keefe, 2002) has been prominent in persuasion research. We shall describe the Summative Model of Attitude in detail in this sub-section. This model has bearing upon the computational approach to persuasive strategies in the proposed artificial agent.

In Figure 2.2.3.1, we show the link from attitude to behavior in a receiver. As put forth by Ajzen (2012), a person’s attitude towards an attitude object determines their behavior towards that attitude object. A person with a positive attitude towards a product (attitude object) will likely engage in the behavior of buying the product over another competing product. A person with a negative attitude towards a political candidate (the attitude object) will engage in the behavior of not voting for that candidate. Under this model, the
communicator engages in the process of persuasion with the intention of changing the attitude of the receiver towards the attitude object by communicating with them, and sending them messages. The messages will be more effective if they are tailored to the recipient, intended to maximize the persuasive impact. In order to change the attitude and ultimately behavior, the summative model provides anchors for tailored messages in beliefs of the receiver that the communicator can utilize to change attitudes.

![Diagram of persuasion process for attitude and behavior change](image)

**Figure 2.2.3.1. Persuasion process for attitude and behavior change**

The Summative Model (Fishbein and Ajzen, 1975) is also known at times as the Expectancy-Value model. The claim in this model is that a person's attitude towards an object is a function of their salient beliefs about the object. For a given attitude object (say, a product, political candidate or an issue), a person may have a large number of beliefs about that object. But at any given time, only certain beliefs are likely to be salient or operational, and those determine the person's attitude towards the object at that time.
The model also postulates belief strength and belief evaluation and gives the below formula:

\[ A_o = \sum b_i e_i \]

where \( A_o \) is the attitude towards the object

- \( b_i \) is the strength of a given belief
- \( e_i \) is the evaluation of given belief

Fishbein and Raven (1962) also proposed procedures for assessing the elements of this model. The following is taken from O’Keefe (2002) to illustrate via example. Suppose that the attitude object being surveyed is Senator Smith. We wish to learn how likely it is that a constituent will vote for the senator. A survey could be constructed that would ask questions regarding the constituents’ beliefs that the senator supports defense cuts, is helpful to constituents, is respected in the Senate and is unethical.

The strength \( (b_i) \) of these belief statements may be surveyed using questions such as the following on a 7-point Likert scale (-3 to +3).

*Senator Smith supports defense cuts.*

very likely

**very unlikely**

The evaluation \( (e_i) \) of these belief statements may be surveyed using questions such as the following on a 7-point Likert scale (-3 to +3).

*Supporting defense cuts is*

very good

**very bad**
The belief strength and evaluation can also be assessed on scales with different Likert scales. For instance, one other scale for measuring belief evaluation has the endpoints *very desirable* and *very undesirable* instead of *very good* and *very bad*. Similarly, one other scale for measuring belief strength would be *true* and *false*, instead of *very likely* and *very unlikely*.

A particular respondent’s responses on these questions in the survey may look like those shown in Table 2.2.3.1.

<table>
<thead>
<tr>
<th>Senator Smith</th>
<th>b₁</th>
<th>e₁</th>
<th>b₁ * e₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>supports defense cuts</td>
<td>+3</td>
<td>-2</td>
<td>-6</td>
</tr>
<tr>
<td>helpful to constituents</td>
<td>-3</td>
<td>+3</td>
<td>-9</td>
</tr>
<tr>
<td>respected in Senate</td>
<td>+2</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>unethical</td>
<td>-2</td>
<td>-3</td>
<td>+6</td>
</tr>
<tr>
<td>∑ b₁ * e₁</td>
<td></td>
<td></td>
<td>-7</td>
</tr>
</tbody>
</table>

Table 2.2.3.1. Survey respondent’s belief strength and evaluation elicited through survey about attitude object Senator Smith

This particular respondent believes that it is quite likely the senator supports defense cuts (belief strength of +3) and supporting defense cuts is seen as a negative characteristic (evaluation of -2). Along the same lines, it is the respondent’s opinion that it is rather unlikely (-2) Senator Smith possesses the negative character trait of being unethical. This hypothetical respondent with a total of -7 (∑ b₁ e₁) can be said to have a negative attitude toward Senator Smith.

The power of the summative model lies in the fact multiple persuasion strategies are possible by how its dimensions are combined; a persuader may try to change the strength of a current belief, change the evaluation of a current belief or indeed change the set of salient beliefs, say by adding a new favorable belief. The model is also helpful in identifying foci for persuasion appeals.
Those beliefs that appear to be contentious or are common ground between opposing views on the same issue can be viable targets for persuasion as discussed in the context of the data shown in Table 2.2.3.2. Consider the following responses (Table 2.2.3.2) by two hypothetical groups of respondents to a survey about nuclear energy, taken and adapted from an example in O’Keefe (2002). Responses are made on 7-point Likert scale (-3 to +3).

<table>
<thead>
<tr>
<th></th>
<th>Pro-Nuclear Energy</th>
<th>Anti-Nuclear Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b_i )</td>
<td>( e_i )</td>
</tr>
<tr>
<td>Prevents future energy crisis</td>
<td>+2.8</td>
<td>+2.7</td>
</tr>
<tr>
<td>Increase risk of nuclear accident</td>
<td>-2.4</td>
<td>-2.8</td>
</tr>
<tr>
<td>Creates waste disposal problems</td>
<td>+2.2</td>
<td>-1.3</td>
</tr>
<tr>
<td>Leads to higher energy costs</td>
<td>+1.9</td>
<td>-2.5</td>
</tr>
<tr>
<td>Leads to less research on renewable energy sources</td>
<td>-1.5</td>
<td>+2.5</td>
</tr>
</tbody>
</table>

Table 2.2.3.2. Survey responses elicited using the Summative Model about Nuclear Energy

The results suggest that preventing an energy crisis is equally favorable to both groups of respondents (mean belief evaluation of +2.7 for both groups) but those who are pro-nuclear energy think that this outcome is more likely (belief strength of +2.8) than those who are opposed to it (belief strength -2.5). A persuasion campaign aimed at inducing favorable attitudes towards nuclear energy would not gain much traction if the messages were aimed at showing how desirable it would be to prevent a future energy crisis. Both groups of respondents already believe that this is a desirable outcome (belief evaluation of +2.7 for both groups). With respect to that belief variable - “prevents future energy crisis” - the anti-nuclear energy group needs to be persuaded not that this is a desirable outcome, but that it is very likely that nuclear power will produce such an outcome. Similarly, messages that focus on providing evidence that developing nuclear energy does not lead to less research on renewable energy sources (belief evaluation is -1.0 and -2.8 for both
groups) would be more successful as an overall strategy aimed at both groups of respondents.

The summative model is thus helpful in defining dimensions (viz., evaluation and strength) along which groups differ and thus identifying multiple viable foci for persuasion appeals. This model and its subsequent refinements detailed in the Theory of Planned Behavior and Theory of Reasoned Action (TRA/TPB) (Azjen, 2012) have been used to devise interventions in a plethora of settings with a highly successful track record. A review of the literature indicates that the theory has been used to successfully design interventions for overcoming speeding behavior (Warner and Åberg, 2008), promoting healthy sexual and reproductive practices (Aarø et al., 2006; Abraham, Henderson and Der, 2004), and promoting exercise and recycling (Abraham and Sheeran, 2004; Aguilar-Luzón et al., 2012), among others.

We exploit the summative model and its elements of belief evaluation and belief strength in an algorithm that is programmed into an artificial agent. To our knowledge, the work undertaken in this thesis is the first to use the summative model for designing an artificial agent, that is, automatic intervention strategies, capable of persuasion. Accordingly, we shall review the literature on artificial agents next.

2.3 Artificial Agents

In the domain of Artificial Intelligence, several approaches have been proposed to tackle the issue of human-machine interaction. Early automated dialogue systems such as ELIZA (Weizenbaum, 1966) and PARRY (Colby, 1981) could conduct a one-on-one “conversation” with a human using rules and pattern-matching algorithms. More recently, the addition of heuristic pattern matching in A.L.I.C.E (Wallace, 2009) led to the development of chat bots
using AIML\(^1\). Many of these systems were inspired by the challenge of the Turing Test (Turing, 1950) or its more limited variants such as Loebner Prize (Shieber, 1993). Bots such as Cleverbot (Saenz, 2010), which participated in 2010 Turing test and was judged to be 59.3% human, compared to 63.3% human achieved by humans, have pushed the limits of chat bot capabilities. The goal in the Turing Test is for a machine to exhibit intelligent behavior, so that a human judge evaluating natural language conversations between a human and the machine would consider the machine to be human. Achieving a rating of 59.3% human means that out of the 334 judges who read transcripts of conversation between a human and Cleverbot, 59.3% judged the transcripts to be between two humans, not human-machine. The Cleverbot has a database of around 265 million rows, containing conversations with humans it has had in the past, from which it learns to respond appropriately to human input. During a conversation, it uses fuzzy matching techniques to compare the entire context of conversation to the millions that it has access to in its database, and responds by finding how a human responded to that input when asked, in part, or in full by Cleverbot in the past. In essence, the Cleverbot is quite effective in sustaining conversation, but it does not have any clear intent to persuade humans. Even more recently, a chatterbot called Eugene Goostman (Veselov, Demchenko and Ulasen, 2014) has been touted to have won\(^2\) the Turing Test. The Eugene Goostman program impersonates a 13-year old Ukranian boy. However, the test was only conducted on 5-minute conversations and impersonating an Ukranian teenager could serve to greatly constrain the conversation topics.

\(^1\)http://www.alicebot.org/aiml.html
With the rise of social media, bots that operate over social media platforms have been created. On micro-blogging platforms such as Twitter, there has been a plethora of bots that scour the tweets of other users and respond in some limited fashion. For example, @tofu_product is a bot that auto-generates tweets after analyzing recent tweets from a user’s timeline. Such capabilities, although at times funny and occasionally on target, do not go beyond very basic capabilities of responding to keywords in other people’s messages. These systems have shallow or no understanding of the ongoing dialogue.

These developments in the field of chat bots that have participated and made strides in the Turing test and social media bots that have garnered followings on sites such as Twitter greatly advance the field of artificial intelligence and human-machine interaction. Nonetheless, one critical drawback of these systems is a lack of agency; these are bots not agents. Franklin and Graesser (1996) defined autonomous agents as “a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.” In the work listed above, the bots do not pursue their own agenda or try to affect their environment. By contrast in this thesis, we created a conversational agent with a very explicit goal of persuasion.

Isbister and Doyle (2002) proposed a taxonomy for organizing the space of embodied conversational agents (CAs). Embodied conversational agents (Nass, Isbister and Lee, 2000; Cassell, 2000; Traum et al., 2012) are different from traditional CAs, due to the introduction of virtual characters with the embodiment of faces, and sometimes even bodies. An animated face in the embodied CA allows the agent to produce non-verbal as well as verbal expressions to engage the user. Accordingly, in our work, we use a narrower
organizing space relative to the taxonomy proposed by Isbister and Doyle (2002) for embodied CAs. We use the organizing space as proposed by Allen et al. (2001), shown in Figure 2.3.1. For each dimension in the taxonomy proposed by Allen et al. (2001), we give examples to delineate the state-of-the-art in the field.

![Figure 2.3.1 Organizing space for conversational agents along dimensions of technique, task, complexity and dialogue phenomena](image)

**Finite state scripts** such as those used by Rosé et al. (2003) are simple representations that move the agent from one state of a finite-state machine to the next depending on the input sequence. These approaches are, however, limited to short interactions and can only be used for the simplest of tasks, such as helping users place long-distance calls. **Frame-based approaches** include most of the spoken dialogue systems constructed to date. In this approach, the system interprets the input to acquire enough information to perform a specific action, be it answering a question about train or flight arrivals or routing a call to the appropriate person at a bank (Zue et al., 2000; Seneff and Polifroni, 2000). The
drawback for such systems is that they are highly specialized and restricted to the specific domain for which they are programmed. At the next level are sets of contexts approaches, which involve representing the user’s task as a series of frame-based contexts. Travel booking agents (Xu and Rudnicky, 2000) and automated bank representatives (Hardy et al., 2004) are examples of sets of contexts approach that can handle complex user interactions. Such systems can perform quite challenging tasks of understanding the user’s goal like making a financial transaction or booking a trip and handling various conversational situations like failure and repairs, using robust data-driven methods (Hardy et al., 2004). These approaches, although adaptable across many different applications, work with structured data to fill ‘slots’, where the data can be flight information or customer account information. Richer representations like the ones used in plan-based approaches (Freedman, 2000; Bohus and Rudnicky, 2003) model the conversational goals of the agent and use planning algorithms to determine a sequence of steps that can achieve the goals. In the plan-based approach, the dialogue involves interactively constructing a plan with the user (for example, a design for a kitchen, a plan to evacuate personnel off an island). Such approaches have been shown to be flexible and robust in conversational situations like mixed-initiative dialog, because the user and agent can work together to achieve the task. The human typically has knowledge of the high-level goals and means of achieving them, while the computer is good at managing the multitude of low-level details that go into such plans. In the plan-based approach, a considerable amount of effort is involved in specifying the goal representations, operators, potential steps, pre-conditions, etc., required by the underlying planning algorithms for such systems. At the last level of complexity in the proposed taxonomy by Allen et al. (2001) are
agent-based approaches. We note here, that in this context, the system components are called agents, which are components assigned with their own tasks and goals. In both the plan-based approach and the agent-based approach, the system is required to maintain an explicit model of task or the world, and to reason about the model. Dialogues with such component agent-based systems involve planning, as well as executing and monitoring a dynamic world, for example, emergency rescue coordination (Blaylock, Allen and Ferguson, 2002). In such CAs, the component agents interactively and collaboratively enable the user to achieve her goals. For example, the system might contain an agent component tasked with finding a plausible interpretation of what the user action or input might be, and another agent component tasked with recognizing the communicative acts of the user input (Burstein, Ferguson and Allen, 2000).

Of the approaches we have reviewed in this taxonomy, the agent-based approach is closest to the work presented in this thesis. In designing our persuasion agent, we created separate software components, each with their own set of tasks – for example, interpreting the input based on speech acts, interpreting based on discourse context, interpreting based on the user model, etc. (McTear, 2002). Notably, one dimension that can be used to draw a contrast between existing designs and our design is the environment in which agent-based systems operate. Traditionally, such systems have been limited to one-on-one situations, where a single agent converses with a human user. There is recent interest in developing CAs operating in multi-party interactive scenarios, with applications including pedagogy (Kumar et al., 2011), social robots for entertainment and companionship (Klotz et al., 2011), and embodied agents for immersive virtual worlds (DeVault and Traum, 2012). Complementary to the goal of creating agents operating in multi-party scenarios, has been
the goal of adding social capabilities to these systems, especially in embodied CAs and robots. Our work in this thesis fits in this space of multi-party, socially competent conversational agents.

Nonetheless, the most important distinction of our work is in the agency of the system. In all the literature reviewed so far, the system does not have a goal independent of the user’s goal. Indeed, the goal of the system is to facilitate the human in achieving her intended task.

The goal of the agent in our work is to persuade the participants during conversation, which may even be in competition with the participants’ own goals. In our case, the agent is required to maintain and update models of the participants’ attitudes and execute the steps necessary to achieve the target attitude. The area of research pertaining to intelligent persuasive systems is relevant here. A number of approaches have investigated automation of persuasive strategies in agents, drawing upon ongoing research in natural argumentation and computational linguistics (Gilbert et al., 2003; Grasso, 2002; Guerini, & Stock, 2005). However, these approaches also require planning systems which prescribe a set of complex steps to generate persuasive dialogue moves and the necessity of identifying arguments made in user’s utterances to respond appropriately. Recent works that have studied applications of natural argumentation include Walton & Reed (2002) and Das (2002). Zukerman et al. (2000) rely on a Bayesian approach for argumentation-based text generation. Gilbert et al. (2003) describe a persuasion machine that can identify arguments in the user’s utterances, evaluate their correctness and present tailored counter-arguments to follow a persuasive dialogue; yet the system they describe is hypothetical.

A persuasive strategy based primarily on argumentation can be considered a resource for persuasion, not the solution. As argued by Guerini, Stock and Zancanaro (2003), persuasion
is a superset of argumentation. They posit that “while argumentation is concerned with the goal of making the receiver believe a certain proposition (goal to induce a belief), persuasion is concerned with the goal of making the receiver perform a certain action (goal to induce an action)”. The link relies on the fact that, apart from coercion, the most efficient way to make someone doing something is to change her beliefs (Castelfranchi, 1996). If our goal is to induce an action then we must also have the goal to induce a belief. In this respect, argumentation is a resource for persuasion.

Therefore, we propose an approach that is:

   a) firmly grounded in social influence research described in Section 2.2;
   b) straightforward to implement, described in Chapter 4;
   c) extensible to any domain, as discussed in Chapter 6; and
   d) empirically validated, detailed in Chapter 5.

In doing so, we combine the sound theoretical underpinnings from social science literature and the best strategies for designing conversational agents from artificial intelligence research.

Another area of research that has bearing upon this thesis is one that deals with computational analysis of discourse, because the agent needs to understand the discourse in order to achieve its persuasion goals. We shall describe some relevant work in the next section.

2.4 Computational Analysis of Discourse

Language-engineering techniques have been crafted in prior research projects (e.g. Carberry & Lambert, 1999; Blaylock, 2002; Strzalkowski et al., 2005; Webb et al., 2005; Small et al., 2009; Shaikh et al., 2010c; Keizer, Petukhova and Bunt, 2011; Bunt, 2011), to
the point that it is now possible to derive models of dialogue occurring in on-line chat rooms and subsequently correlate these with types of behavior identified in small group dynamics research. Dialogue has been studied from a variety of different perspectives, including computational modeling of dialogue acts. Dialogue Acts (Bunt, 1994) are very much related to Speech Acts introduced by Austin (1962) where each utterance can be explained by understanding the action that the speaker is attempting to accomplish. Another critical piece of prior research we shall utilize in our work is the robust detection of sociolinguistic phenomena in multi-party discourse (Strzalkowski et al., 2010; Shaikh et al. 2012; Broadwell et al., 2013) Using a two-tier approach of discovering linguistic features from conversational text to inform models of sociolinguistic behavior such as Topic Control, Disagreement and Involvement, an automated model of Influence, Leadership and other social phenomena has been built. Our counter-persuasion strategies depend upon the key insights drawn from correlations of influence with its sociolinguistic behaviors.

2.4.1 Dialogue Act Tagging

The responses made by the agent are triggered, in part, by the function of the utterance the agent is responding to. Our objective is to capture how an utterance functions in dialogue, which may or may not be directly related to its form. This means that we want to know if a particular utterance is functioning as a question, a confirmation or a statement of facts, its syntactic construction notwithstanding. Statements, questions, answers, offers, acceptances and rejections, as well as expressions of thanks are all examples of such functions in a dialogue, which are called Dialogue Acts (DA). For example, the utterance “Can you close the window?” may function as a question or as a directive, depending upon the context in which it is used.
Our taxonomy of dialogue acts derives from prior work that we have undertaken in discourse modeling (Shaikh et al., 2010b). The Figure 2.4.1.1 shows the hierarchy of dialogue acts we have adopted in this thesis. Dialogue Acts are classified into three hierarchical categories: (a) Statements-and-Responses, (b) Questions-and-Directives, and (c) Conversational-Norms. Each of these categories consists of several top-level tags and also contains specialized tags under these.

A Statement makes a claim about the world, and tries to change the beliefs of the listener. In general, an utterance that is a statement can be said to be true or false. Responses are utterances that users make to indicate reaction to another user’s utterances, such as answering it, acknowledging it or agreeing to it. Agree-Accept tags are applied when the
user accepts or agrees with another user’s proposal or request; or if the information or claim conveyed in a statement is accepted or confirmed. Similarly, when the user disagrees, rejects a proposal or offer, says he will not comply, or says that the claim or the information expressed by the other user is incorrect, we use the Disagree-Reject tag.

Questions and directives are typically those that elicit some response from other users. For example, an Action-Directive places an obligation upon the listener to respond: either to perform the action as requested or to refuse to do it. An Action-Directive would often be followed by either Agree-Accept or Disagree-Reject utterances. The group of tags under Conversational Norms category deals with the social norms occurring in conversations. A complete description of the tags used in this work is presented in Appendix F.

We use this hierarchy of dialogue acts to tag utterances for the agent to base its behavior upon and use an automated classifier to do so. Some models of dialogue act assignment require complex planning and belief-based interpretation, such as the work of Cohen and Perrault (1979), Perrault and Allen (1980) and Lee and Wilks (1996). Statistical modeling and machine learning techniques such as neural networks (Ries, 1999), Hidden Markov Models (Stolcke et al., 2000), LSA (Serafin and Eugenio, 2004), as well as Bayesian approaches (Grau et al., 2004) have also been employed. In these approaches, existing corpora of annotated data are used to train classifiers. The properties of utterances in these annotated corpora are exploited – such as n-grams and dialogue act tags of prior utterances, and in the case of spoken data, prosodic features.

In our work, we adapt the cue-based dialogue act tagging approach developed by Webb (2010). In this approach, words or phrases are used to indicate the dialogue acts of the utterances, serving as reliable indicators of the utterance discourse function. Our primary
reason for using this approach over others is its reliability, ease of use and speed of using such an implementation to tag utterances in real-time conversation. As part of our assessing the current state-of-the-art in dialogue act tagging, we did a comparison of real-time response time in automatically tagging utterances and the cue-based approach provided the most promising results in our experiments in terms of optimizing prediction accuracy and response lag time. The agent uses the automatically assigned dialogue act tags to understand the function of participants’ utterances during conversation and respond accordingly.

2.4.2 Modeling Sociolinguistic Phenomena in Discourse

Another critical piece of prior research we utilize in this work is the robust detection of socio-linguistic phenomena in multi-party discourse (Strzalkowski et al., 2010; Shaikh et al., 2012; Broadwell et al., 2013). In this work, a two-tiered approach is used to model social phenomena in conversations, such as leadership, influence, pursuit of power and group cohesion. The first tier combines lower-level linguistic features such as local topics, communication links and dialogue act in order to determine a range of sociolinguistic behaviors that participants display in conversation, including topic control, task control, disagreement, network centrality, among others. At the second tier, these behaviors are combined to rank participants by their degree of higher-level sociolinguistic phenomena in the group, such as their leadership or influence. This method is described in detail in Broadwell et al. (2013) and has been shown to have a high accuracy (90%) across many forms of interaction and across many languages (Taylor et al., 2012).
In Shaikh et al. (2012), we have elaborated upon the two-tiered influence detection algorithm. We shall briefly explain here a subset of elements from that work that has bearing upon the agent algorithm.

For example, the *Topic Control Measure (TCM)* is computed using several indices, two of which are the Local Topic Introduction Index and the Subsequent Mention Index, described below. In any conversation, whether it is focused on a particular issue or task or is just a social conversation, the participants continuously introduce multiple topics and subtopics. These are called *local topics*. Local topics, following the notion put forth by Givón (1983), may be equated with any substantive noun phrases introduced into discourse that are subsequently mentioned again via repetition, synonym, or pronoun. Some of these local topics may be talked about in only a couple of turns, while others may persist for much longer; some of them will be relevant to the overall discussion, while others may appear like digressions. Who introduces local topics into conversation and who continues to talk about them, and for how long are some of the indicators of topic control in dialogue.

*Local Topic Introductions (LTI):* Participants who introduce more local topics exert more topic control in dialogue (Palmer, 1989; Rienks et al., 2006). This index calculates the proportion of local topics introduced by each participant, by counting the number of first mentions of local topics by each participant as percentage of all local topics in a discourse.

*Subsequent Mentions of Local Topics (SMT):* Participants who introduce local topics that are subsequently widely discussed exert a high degree of topic control in discourse (Palmer, 1989; Rienks et al. 2006). This index calculates the percentage of discourse utterances where the local topics introduced by each participant are being mentioned (by themselves or others) through repetition, synonym, or pronoun.
When we compute the correlation of each index that constitutes Topic Control with the behavior, we see that the Local Topic Introductions (LTI) and the Subsequent Mentions of Local Topics (SMT) correlate very highly with Topic Control Measure (TCM). In Table 2.4.2.1, we show the correlations obtained from a sample online chat dialogue. We see that the Local Topic Introduction Index and Subsequent Mentions correlate very strongly with the behavior we wish to model – Topic Control (TCM). Another index such as Cite Score (CS), which is a measure of how frequently others cite topics introduced by a participant, excluding self-mentions, also correlates very strongly. The Turn Length (TL) index, which measures the average length of turns for each participant has a high correlation, but not as high as the other indices. The correlation among indices of Topic Control follows a similar trend across a number of discussions. Thus, we can postulate that introducing topics that others subsequently talk about in discourse is more indicative of Topic Control behavior than having utterances that have longer length. We will make use of this finding in our agent behavior model, specifically the counter-persuasion algorithm. Additionally the correlation between sociolinguistic behaviors used to compute influence scores of participants is shown in Table 2.4.2.2.

<table>
<thead>
<tr>
<th></th>
<th>LTI</th>
<th>SMT</th>
<th>CS</th>
<th>TL</th>
<th>TCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTI</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT</td>
<td>0.96</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>0.88</td>
<td>0.85</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL</td>
<td>0.78</td>
<td>0.79</td>
<td>0.80</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>TCM</td>
<td>0.92</td>
<td>0.95</td>
<td>0.93</td>
<td>0.89</td>
<td>1.0</td>
</tr>
<tr>
<td>α</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4.2.1. Correlations among selected Topic Control indices for a sample online chat dialogue

<table>
<thead>
<tr>
<th></th>
<th>NCM</th>
<th>MAD</th>
<th>TCM</th>
<th>CDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCM</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>0.86</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCM</td>
<td>0.98</td>
<td>0.86</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>CDM</td>
<td>0.58</td>
<td>0.59</td>
<td>0.48</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 2.4.2.2. Correlations among influence measures for a sample online chat dialogue
We see that the Topic Control Measure (TCM) shows the highest correlation with Network Centrality Measure (NCM). The NCM calculates the degree to which a participant is a “center hub” of the communication within the group. In other words, the influencer is someone whom most others direct their comments and who engages others in the conversation. The Measure of Argument Diversity (MAD) also correlates strongly with the other measures. MAD can be signaled by a wider vocabulary, including citations of authoritative sources as well as use of specialized terminology (Huffaker, 2010). A person who uses more varied vocabulary and introduces more unique words into a conversation is considered to have a higher degree of Argument Diversity. We note that the Cumulative Disagreement Measure (CDM) has lower correlations with other measures. Disagreement is a part of influence in that it is possible that an influential person in a small group engages in disagreements with others in order to control the topic by way of identifying or correcting what they see as a problem (Sanders, Pomerantz & Stromer-Galley, 2010).

Using insights learnt from these correlations, we can rank the measures – TCM, NCM, MAD and CDM, in order of their importance in the overall influence model. We can also compute the correlation of each behavior with the influence scores of participants. The ground truth for influence scores was again obtained from human judgments collected using post-test surveys (described in Shaikh et al., 2010a). We found that, on average, TCM measure correlates higher than others with survey ratings. As shown in Table 2.4.2.3, the measures for a sample online chat dialogue and their correlations with human survey ratings reveal the relative importance of the behaviors with respect to the influence phenomena. We find that the correlations of CDM are relatively low in chat room type of synchronous
discussions. However, a similar analysis of data from online asynchronous discussion forums showed that CDM has the highest correlation with influence in those settings.

In addition, we computed the coefficient of determination ($R^2$), which is the proportion of variance in the dependent variable (influence) predicted from the independent variables (TCM, NCM, MAD and CDM in this case). In Table 2.4.2.4, we show the $R^2$ values for each of the behaviors with respect to influence. We find that the co-efficient of determination for Topic Control Measure (TCM) and Cumulative Disagreement Measure (CDM) are the highest, followed by Network Centrality Measure (NCM) and Measure of Argument Diversity (MAD). An $R^2$ of 0.85 means that 85% of the variance in influence is predictable from TCM. This is also reflected in the way Shaikh et al. (2012) assign weights to a weighted influencer model.

<table>
<thead>
<tr>
<th></th>
<th>NCM</th>
<th>MAD</th>
<th>TCM</th>
<th>CDM</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCM</td>
<td>1</td>
<td>0.85</td>
<td>0.98</td>
<td>0.58</td>
<td>0.89</td>
</tr>
<tr>
<td>MAD</td>
<td>1</td>
<td>0.85</td>
<td>0.59</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>TCM</td>
<td></td>
<td></td>
<td>0.48</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>CDM</td>
<td></td>
<td></td>
<td></td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.4.2.3. Correlations of behaviors with human survey ratings

<table>
<thead>
<tr>
<th></th>
<th>TCM</th>
<th>NCM</th>
<th>MAD</th>
<th>CDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.85</td>
<td>0.63</td>
<td>0.51</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 2.4.2.4. Co-efficient of determination $R^2$ for each measure with influence

We leverage the important findings from this section to automate the persuasion and counter-persuasion strategy in our agent. Our counter-persuasion strategy is built around the most important behaviors indicative of influence – Topic Control and Disagreement. Additionally, we will use the previous result that introducing topics that others talk about in conversation are the more important for Topic Control behavior than having longer
turns for automating Topic Control in the agent. We explain the implementation of counter-persuasion strategies in Chapters 4 and 5.

We have reviewed the discourse analysis literature; next we turn our attention to the issue of creating representations of mental states of the participants. We want to characterize the participant’s attitudes towards topics of discussions in terms of their belief strength and evaluation, as necessitated by the summative model of attitude we have chosen. To do so, we use the Topical Positioning method described in the next section.

2.5 Topical Positioning

Topical Positioning is defined as the attitude a speaker has towards the meso-topics of discussion. Meso-topics are those topics that stretch for a length of discourse, and participants tend to make polarized statements about these topics. Participants in a dialogue, when discussing issues, especially ones with some controversy, signal their attitude on each topic, classified as for, against, or neutral/undecided. In doing so, they establish their positions on the issue or topic. The participants’ positions then shape the agenda of the discussion and also shape the outcomes or conclusions of the discussion. Determining topical positioning allows us to designate the speakers as for, against, and neutral/undecided on a given topic or issue.

In prior research, we have worked on Topical Positioning in conversation (Lin et al, 2013). To establish topical positioning, Lin et al. (2013) first identify meso-topics that are present in a discourse. Next, the polarity of each utterance containing a meso-topic is determined, i.e., if this utterance is ‘for’ (positive) or ‘against’ (negative), or neutral on a meso-topic. Topical Positioning is then represented by a single number that is based on two component measures: a Topic Polarity Index, which establishes the polarity of a speaker’s attitude
towards the topic, and a Polarity Strength Index, which measures the magnitude of this attitude.

[Topic Polarity Index (TPX)] In order to detect the polarity of Topical Positioning on meso-topic T, for each speaker the following are counted:

1) All utterances on T using statements with polarity P applied directly to T using appropriate adverb or adjective phrases, or when T is a direct object of a verb.

2) All responses to other speakers’ statements about T with polarity P.

The polarities are calculated using the valence scores of words from the expanded ANEW lexicon (Bradley and Lang, 1999; Shaikh et al., 2016). The ANEW (Affective Norms of Words) lexicon (Bradley and Lang, 1999) is a lexicon containing valence scores of approximately 2200 words ranked on a discrete scale from 1 (most negative) to 9 (most positive). We (Shaikh et al., 2016) expanded this lexicon to over 100,000 words by adopting the method developed by Liu et al. (2014) in expanding the MRC imageability lexicon. While other sources of valence judgments exist such as NRC (Mohammad, Kiritchenko and Zhu, 2013) and MPQA (Wiebe and Cardie, 2005), they have serious limitations – for instance – the NRC lexicon rates each words on a positive or negative scale, which does not allow for more fine-grained analysis of strength of valence.

Given the counts of utterances with polarity P (as explained in 1 and 2 above), Lin et al. calculate TPX for each speaker as a proportion of positive, negative and neutral polarity utterances made by this speaker about topic T. A speaker whose utterances are overwhelmingly positive (80% or more) has a pro-topic position (TPX = +1); a speaker whose utterances are overwhelmingly negative takes an against-topic position (TPX = -1);
a speaker whose utterances are largely neutral or whose utterances vary in polarity, has a neutral/undecided position on the topic (TPX = 0).

[Polarity Strength Index (PSX)] In addition to the valence of the Topical Positioning, Lin et al. also calculate its strength. To do so, they calculate the proportion of utterances on the topic made by each speaker to all utterances made about this topic by all speakers in the discourse. Speakers, who make most utterances on the topic relative to other speakers, take a stronger position on this topic. PSX is measured on a 5-point scale corresponding to the quintiles in normal distribution.

Topical Positioning Measure (TPM)

To calculate the value of Topical Positioning for a given topic for a participant, the values of TPX and PSX are combined by taking a product (TPX*PSX). Topical Positioning takes values between +5 (strongest pro) to 0 (neutral/undecided) to −5 (strongest against). For example, a participant who makes 25% of all utterances on the topic “Carla” (group mean is 12%) and whose most statements are positive, has the strongest pro Topical Positioning on Carla: +5 (for fifth quintile on the positive side). In Table 2.5.1, we show the TPX and PSX scores for a participant named WB from Lin et al.’s work. The scores are shown for a set of five meso-topics (cops, drunk driving, etc.) related to the discussion topic – whether the minimum legal drinking age should be lowered from 21 to 18. In Table 2.5.2, we show the corresponding Topical Positioning vector, which contains the product of TPX and PSX. Such a vector establishes participant WB’s position on the overall topic of discussion. We can see that WB is positive towards meso-topic cops, but negative towards the rest of the meso-topics, and using this information her stance on the overall topic may be construed as negative.
Lin et al. have experimentally shown an accuracy of 87% in predicting changes of opinion of participants in multi-party online conversations using the topical positioning measure. The algorithm performance was validated against ground-truth data obtained from responses on pre- and post-discussion surveys, which characterized participants’ opinions on the topic of discussion.

When we consider the indices TPX and PSX for the Topical Positioning method (as shown in Table 2.5.1), and the belief strength and evaluation \( (b_i \text{ and } e_i) \) as proposed by the Summative Model (Section 2.2), the parallels are striking. We decided to adapt the topical positioning method to model the beliefs of participants during conversation and ascertain the values of belief strength and evaluation as a multi-dimensional matrix. We will explain the adaptation we made in detail in Chapter 4.

### 2.6 Summary of this chapter

In this chapter, we started with a review of literature pertinent to persuasion research. We provided the motivation for choosing the Summative Model of Attitude as a basis for programming the persuasion strategies in the agent. Next, we reviewed the state-of-the-art in artificial agents and their capabilities. Agents that could achieve social goals of the kind we propose in this work were found to be wanting in prior research in this field. In the
domain of conversation analysis, we reviewed literature on dialogue act tagging and modeling social phenomena and found relevant scholarship that can be carried towards new ground by our proposed work. The capability of automatically identifying dialogue act tags is an important component in the agent architecture, since the dialogue act tags of participant utterances drive the reactions of the agent. We next reviewed modeling sociocultural phenomena in discourse and argued in favor of using Topic Control and Disagreement behaviors as the basis for programming counter-persuasion strategies in the agent. Finally, we examined the Topical Positioning method of modeling participant behavior and detecting opinion change in discourse and found them to be sufficiently advanced and relevant to our work, so that it can be directly tailored to suit our needs. In Chapter 4, we propose a consolidated approach that effectively makes use of this varied body of work. But first in Chapter 3, we describe a belief elicitation study that was undertaken as an essential step towards realizing our overall research goal.
Chapter 3: Belief Elicitation Study

3.1 Introduction

Our goal is to use the Summative Model, as described in the Literature Review Section 2.2, to design an agent that can deliver interventions that change attitudes in ongoing conversation. The Summative Model, as well as its more enhanced versions in the Theory of Reasoned Action/Theory of Planned Behavior (TRA/TPB) (Ajzen, 2006a), suggests that a person's attitudinal, normative and control beliefs toward a given behavior are the foundations on which intentions to perform, and actual performance of behavior are pinned. Elicitation studies are typically used to identify salient beliefs individuals hold toward a given behavior (Ajzen, 2006b). Despite the importance of the belief elicitation stage towards intervention development with respect to TRA/TPB, to date, little belief elicitation research has been published (Sutton et al., 2003). A rigorous review has revealed that research on elicitation studies to design interventions for socially relevant discussion topics is absent in the literature, making this study possibly the first of its kind.

With respect to language research, there has been ongoing interest in investigating the extraction of beliefs from text, as well as, understanding, modeling, and ascription of beliefs to the speaker (Ballim, Wilks and Barnden, 1991; Diab et al., 2009; Werner et al., 2015). In NLP, the terms “belief” and “factuality” are sometimes used to refer to the same phenomenon. Under this framework, language research is focused on whether the communicator firmly believes a statement she is making to be true or weakly believes it to be true (Prabhakaran et al., 2015). This leads to the possibility of interpreting beliefs as
being hedged: “The ball is sort of blue.” or “The ball may be blue”; or modal beliefs – whether or not the communicator thinks something to be true versus whether she thinks something to be untrue (“I don’t think Trump will win the election.”). For example, Diab et al. (2009) and Prabhakaran, Rambow and Diab (2010), classify propositions in text as beliefs in three categories –

(a) committed beliefs: the writer strongly believes an assertion to be true;
(b) non-committed beliefs: the writer probably or possibly believes an assertion to be true, but is not certain; and
(c) not applicable: the writer expresses some other cognitive state about the assertion, such as desire or intention, or expressly states that she has no belief about the proposition by asking a question.

Another category of (d) reported beliefs (where a writer attributes beliefs to another person or group of persons) has been added to this list of categories recently (Prabhakaran et al., 2015).

For our purposes though, we describe beliefs as the state of mind of an individual who thinks something to be the case, regardless of whether there is factual evidence to support such belief. In our framework, beliefs are propositions that are held true by a belief source about a belief target. In these propositions, we situate the belief targets as occupying certain semantic roles. We move beyond linguistic markers of the degree of commitment held towards a belief by the communicator or the presence of hedge words in the sentence towards a deeper understanding of the cognitive state of the communicator. We do so by incorporating both the strength of the belief and the evaluation placed towards the belief into a unified construct. Additionally, we use the summative model to tie the beliefs directly
to attitudes held by the communicator, which in turn allows us a handle on their intent to perform behaviors relative to their underlying beliefs.

As mentioned above, while belief elicitation studies have been conducted to gather data regarding target behaviors, a thorough review of literature suggests that our study is the first to create a corpus where attitude towards socially relevant topics is directly linked to and can be ascertained from the underlying beliefs. In our corpus, the attitude objects are topics or concepts, such as *Genetically Modified Foods* or *Lowering drinking age from 21 to 18*; not behaviors to be performed, like exercising or dieting.

In the following sections of this chapter, we describe our motivation to undertake this elicitation study. We then describe in detail our experimental procedure to collect the data. Next, we describe the analysis of data, along with related descriptive and inferential statistics.

### 3.2 Motivation

Our goal in creating this corpus was primarily to model underlying attitudes that individuals hold towards socially relevant discussion topics. The data can also be used by many in the NLP community who are interested in research related to beliefs, and also sentiment and opinion modeling.

Belief elicitation studies are typically conducted by researchers in the social sciences who wish to gauge the salient beliefs of a population towards a particular attitude object. Typically, health-related behavioral interventions are designed using elicitation studies (Manstead and Parker, 1995; Sutton et al., 2003, Dean et al., 2006). For our purposes, targets of attitude are “topics” – which may include objects or people (e.g. political candidates, commercial products), social issues (e.g. living together before marriage,
divorce) as well as activities and behaviors (e.g. smoking, exercising, condom use). Our study follows the procedures suggested by Ajzen & Fishbein (1980) to elicit beliefs in a sample population.

We have two distinct reasons to undertake this study. The first motivation is to use the beliefs so elicited to design survey instruments. The surveys are to be administered to participants before and after our discussion sessions with the agent, which are the *interventions*. These surveys are crucial to our work, because the responses on the surveys will provide the ground truth from which to measure the agent’s efficacy in being persuasive. Using the salient beliefs elicited in this study to design the surveys makes it more likely that the questions on the survey will tap into the salient beliefs of the participants of the discussion sessions.

Our second motivation is to use the corpus of beliefs elicited as a database of natural language statements for the agent in the discussion. We have made the case in previous chapters that we do not wish to undertake the task of natural language generation for the agent. However, the agent needs to communicate with the other participants in order to persuade them and this corpus circumvents the need for natural language generation. Our solution to the problem of natural language generation for the agent is to use the responses made by humans on the topics as a database of arguments made in favor of and against the topic of discussion. We will pre-program these arguments in the agent, with the intention of using them at opportune moments in the conversation. This procedure is described in Chapter 4, Section 4.5, Behavior Selector.
As noted, we will use the data generated from this study in two different ways to further our goals of creating and testing the persuasive agent. We now describe the experimental procedure of collecting salient beliefs, beginning with the description of participants.

3.3 Participants

Our sample consisted of 67 individuals recruited from the undergraduate student population at our university, of which 56 (83.5%) completed the study in its entirety. We collected information about their age, gender, ethnicity and degree of major at the university. We also collected their email addresses in order to send follow up requests.

The average age of participants was 19 years (sd=1.17). The sample consisted of 32 respondents identifying as female and 24 as male (1 respondent preferred not to answer). Table 3.3.1 shows the demographic characteristics of our recruitment sample, those who completed the study and those who did not.

Representativeness Check: We conducted chi-square analysis to check the representativeness of our sample and found no significant differences with respect to gender ($\chi^2(1)=.5041$, $p=.47$) or age ($\chi^2(1)=1.3723$, $p=.50$) at significance levels of $p < 0.05$ between those who participated in our study and the target population.

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Whole sample (n=67)</th>
<th>Completers (n=56)</th>
<th>Non-completers (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>19</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>SD</td>
<td>1.3</td>
<td>1.2</td>
<td>1.28</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (%)</td>
<td>54%</td>
<td>57%</td>
<td>37%</td>
</tr>
<tr>
<td>Male (%)</td>
<td>46%</td>
<td>43%</td>
<td>63%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian (%)</td>
<td>67%</td>
<td>71%</td>
<td>83%</td>
</tr>
<tr>
<td>Non-Caucasian (%)</td>
<td>33%</td>
<td>29%</td>
<td>17%</td>
</tr>
</tbody>
</table>

Table 3.3.1. Demographic comparisons of participants who completed the study and those who did not
3.4 Topics

Our goal was to elicit the salient beliefs from our sample population regarding topics or issues that were of social relevance and importance to them. The six topics we chose for our study are:

1. Genetically modified foods
2. Living together before marriage
3. Being able to easily identify people online
4. Easy access to morning after pill to people under 17
5. Lowering minimum legal drinking age to 18 from 21
6. Stricter airport screening procedures

These topics were chosen based on an extensive survey of relevant topics found on the internet across various discussion forums, online debate sites, news articles and opinion pieces, and blog articles. We narrowed our selection to the six topics upon careful consideration of several factors, including the below:

1. Whether participants would be relatively well informed about the topic
2. Whether participants would have prior opinions about the topic
3. Whether participants would be interested in discussing the topic at length for a period of time.

3.5 Survey Questions

According to Francis et al. (2004), in order to predict whether an individual intends to perform a behavior, viewed as the proximal determinant of actual behavior (Ajzen, 1991), a researcher needs to know the beliefs of the individual toward the behavior under consideration.
Accordingly, we used the recommendation from prior research (Ajzen & Fishbein, 1980) and framed our open-ended questions as below:

1. What do you think are the advantages of $X$?

2. What do you think are the disadvantages of $X$?

In both these questions, $X$ would be replaced by one of the six topics listed in Section 3.4 above. In Figure 3.5.1, we show a screen capture of the questions on the survey regarding 

*Topic 1: Genetically modified foods*. Participants were instructed to carefully consider their responses and write them in free-form natural language. There was no upper limit on the time allowed or number of characters allowed for the responses.

![Figure 3.5.1. Elicitation questions on one of topics in an actual screen capture](image)

### 3.6 Procedure

Participants were recruited via email and in-class announcements. They were asked to respond to surveys, which were administered online, hosted on a secure server maintained
by the research team. Ethical approval was obtained from the Institutional Review Board before data collection commenced. Online elicitation was chosen for two main reasons. First, traditional one-to-one interviews and focus groups techniques can take considerable time to organize, undertake, and transcribe (Sommer and Sommer, 1997). Second, traditional interview techniques are prone to bias where participants may respond in socially acceptable ways to please the researcher (Dyer, 1995). The use of anonymous online survey techniques may help to reduce this bias.

The data collection took place over a period of six weeks from late March through early May, 2015. We conducted the study in three parts using three surveys that each contained open-ended questions on only two of the six topics noted in Section 3.4, in order to ensure timely and thoughtful responses from the participants on each topic.

Participants were sent a link to one survey at a time and asked to complete their responses within two weeks. When participants clicked on the link to the survey they were presented with the participant information sheet and consent procedure, before being given survey instructions. We sent two email reminders to those who did not complete the survey, at even intervals within the two-week timeframe for each survey. Once participants completed their responses on one survey, they were sent the link to the next survey, containing the ADVANTAGES and DISADVANTAGES question on two other topics out of the six listed in Section 3.4. In addition to spacing out the surveys to counter response fatigue, we also wanted to ensure that the order of presenting topics and also the questions did not influence the responses given by participants.

In order to do so, we varied the order of topics and questions within the surveys, so that an equal proportion of respondents would be presented with one topic before the other (and
vice versa) and within the topics, would be presented the ADVANTAGES question before the DISADVANTAGES question (and vice versa). Accordingly, there were eight different versions of each survey to account for all possible permutations of topic and question order. Participants were randomly assigned to one possible permutation for each of the eight different survey variations at the outset of data collection. As shown in Figure 3.3.1, the participants who were randomly assigned to one of the survey conditions were given the “What do you think are the disadvantages of Genetically Modified Foods?” question before the “What do you think are the advantages of Genetically Modified Foods?” whereas other participants where shown the questions in the reverse order. Appendix A presents a copy of one version of this survey containing two out of six topics we have chosen.

In Table 3.6.1, we present three sample responses obtained for the topic Genetically Modified Foods from participants. We should note that these responses are obtained from participants who were assigned different versions of the surveys. As can be seen in Table 3.6.1, participants vary with respect to content in their responses on each question with the types of statements and arguments made in response to each question. These responses are the most accessible beliefs that come to mind for these participants when faced with the corresponding question and are taken to be the salient beliefs with respect to this issue. The set of salient beliefs of a population on this issue is then comprised of the set of all such statements and beliefs expressed by our experiment population.

Having collected such data for all six topics, a series of qualitative and quantitative analyses were conducted on the elicitation data. We shall describe our qualitative analysis procedure first in the next section.
<table>
<thead>
<tr>
<th>What do you think are the <strong>ADVANTAGES</strong> of GMF?</th>
<th>What do you think are the <strong>DISADVANTAGES</strong> of GMF?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetically modified foods makes it easier for farmers to grow more crops. This will help solve hunger issues around the world. Also, these would not need any pesticides so people will not be eating any pesticides or any other disease/germs.</td>
<td>The disadvantages is the people who makes the new genetically modified food will have a monopoly. Small farmers cannot compete against them so the farmers will give in and pay for their product. It will cost more for farmers to grow crops. Even though they will make more crops, the selling prices of the crops will go down due to increase of supply. Also, some people believe it is wrong to play God and change up the genes of these food.</td>
</tr>
<tr>
<td>The advantages of genetically modified foods are they are able to decrease the costs of foods and reduce the number of those affected by starvation around the world. Decreasing the costs of GMF’s will allow those in poverty to be able to afford food which leads to the decrease in starvation and the ability for people to save money and provide for their families.</td>
<td>The disadvantages of genetically modified foods are there are unknown side affects towards the health of humans. This can be incredibly dangerous since these side effects can lead to long term, negative health issues and may even lead to death. If the side effects of GMF’s are this harmful it is scary to think that these types of food would be open to the public.</td>
</tr>
<tr>
<td>1) Create more crops  2) More crops could stop starvation around the world.  3) GMF could have more nutrition and flavors.</td>
<td>1) It could be unhealthy because of the amount of chemicals that will be used.  2) It could cause cancer because the engineered material can increase the speed of the tumors growth.</td>
</tr>
</tbody>
</table>

Table 3.6.1. Three separate responses obtained from participants on the topic of Genetically Modified Foods (GMF)

### 3.7 Analysis of Data – Annotation Procedure

Here we describe our annotation procedures to categorize beliefs that are identical but expressed in different ways into broader categories, or *themes*.

The method recommended to analyze elicitation data is to do content analysis (Ajzen 2006b; Francis et al., 2004). Qualitative studies exploring beliefs typically adopt a top-down approach to this analysis, meaning that the theoretical underpinning guides the analysis rather than the data as in a bottom-up approach (Bayley, Brown and Wallace 2009; French et al. 2005; Patch, Tapsell and Williams 2005). Once themes are generated, as explained below, content analysis enables the data to also be explored quantitatively.
(Weber, 1990; Neuendorf, 2002; Sutton et al. 2003; French et al., 2005; Krippendorff, 2012). We adopt this approach in our study as well.

The top-down approach is a three-step process, repeated for all six topics. First, all responses associated with the topic were read multiple times to gain an overview of the beliefs expressed for each issue. Next, we identified the agent, patient and activity in each response; typically these are the subject, object and verb of a sentence. This annotation was conducted at the sentence level, or clause level, if there were multiple beliefs present in a sentence. The third step was to create a broad category where all beliefs that had the same agent, patient and activity could be assigned.

<table>
<thead>
<tr>
<th>Response #</th>
<th>What do you think are the ADVANTAGES of GMF?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Genetically modified foods makes it easier for farmers to grow more crops. This will help solve hunger issues around the world. Also, these would not need any pesticides so people will not be eating any pesticides or any other disease/germs.</td>
</tr>
<tr>
<td>2</td>
<td>The advantages of genetically modified foods are they are able to decrease the costs of foods and reduce the number of those affected by starvation around the world. Decreasing the costs of GMF's will allow those in poverty to be able to afford food which leads to the decrease in starvation and the ability for people to save money and provide for their families.</td>
</tr>
</tbody>
</table>
| 3          | 1) Create more crops  
2) More crops could stop starvation around the world.  
3) GMF could have more nutrition and flavors. |

Table 3.7.1. An example of analysis of beliefs elicited on the topic of Genetically Modified Foods (GMF)

Consider the responses we have shown in Table 3.6.1 for the question “What do you think are the ADVANTAGES of Genetically Modified Foods?” replicated in Table 3.7.1 here. The phrases highlighted in red are some of the beliefs that have been annotated from these responses as part of step two of our content analysis process.
In each of these annotated beliefs, we can postulate that GMFs are the agent and starvation is the patient. In step 3, we manually each of these beliefs to a broad theme:

*Genetically modified foods reduce starvation.*

The themes are created based on beliefs that shared commonality (Krippendorff, 2012). In this example, synonymous phrases hunger issues and starvation are assigned to the same theme manually.

Given the framing of our survey questions (c.f. Section 3.5), the topic or issue was in the agentive role in most of beliefs elicited. In other cases, the beliefs elicited were propertive – in that, they described a property or characteristic of the topic. For illustration, consider the part of response 3 in Table 3.7.1 highlighted in green. From this response and others that expressed the same belief the below theme was generated:

*GMFs contain more nutrients.*

In this manner, all responses were analyzed for each of the six topics and themes were generated in a top-down fashion. All responses and the beliefs contained in them were categorized into the themes so generated.

In order to reduce bias and address credibility, we asked a trained annotator to independently assign the themes generated by the expert to responses. The annotator was given the responses, without any annotation, as seen in Table 3.6.1. The annotator was also given a list of themes that were generated for each topic and were asked whether she would wish to add any more themes to the list.

The annotator suggested rewording a few themes to make them more inclusive. For example, it was suggested that the theme

*Lowering minimum legal drinking age would lead to increase in college dropouts.*
be changed to

*Lowering minimum legal drinking age would lead to poor performance in school.*

The annotator then assigned themes to each of the responses as explained in the foregoing. We computed the inter-annotator agreement between the trained annotator and our own assignment of themes to responses. The value of Cohen's kappa (1960) for the agreement was 0.725, which is above the acceptable rate. Given this acceptable level of agreement, we decided to use our annotations as the basis going forward. In doing so, we follow Sutton et al. (2003), who have previously opted to base their analysis on one annotator's judgment when they found an acceptable level of agreement between annotators.

In the next section, we present descriptive statistics related to the annotated data.

### 3.8 Analysis of Data – Descriptive Statistics

In this section, we provide details about the set of beliefs elicited for all six topics across all respondents. In Table 3.8.1, we show the total number of themes for each topic. In Table 3.8.2, we show the most frequently expressed theme for each topic and the count of number of beliefs that were categorized under those themes.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Total number of themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Genetically modified foods</td>
<td>21</td>
</tr>
<tr>
<td>2. Living together before marriage</td>
<td>14</td>
</tr>
<tr>
<td>3. Being able to easily identify people online</td>
<td>15</td>
</tr>
<tr>
<td>4. Easy access to morning after pill to people under 17</td>
<td>15</td>
</tr>
<tr>
<td>5. Lowering minimum legal drinking age to 18 from 21</td>
<td>16</td>
</tr>
<tr>
<td>6. Stricter airport screening procedures</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 3.8.1. Total number of themes generated for each of the six topics

This shows there were 44 responses that contained a belief categorized under the theme:

*GMFs have a harmful effect on health*

and similarly, 43 responses contained a belief that was categorized as

*Stricter airport screening procedures would lead to increased safety*
<table>
<thead>
<tr>
<th>Topic</th>
<th>Most frequent theme</th>
<th>Number of beliefs in most frequent theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Genetically modified foods</td>
<td>Have a harmful effect on health</td>
<td>44</td>
</tr>
<tr>
<td>2. Living together before marriage</td>
<td>Leads to knowing partner on a deeper level</td>
<td>39</td>
</tr>
<tr>
<td>3. Being able to easily identify people online</td>
<td>Reduces freedom of expression</td>
<td>28</td>
</tr>
<tr>
<td>4. Easy access to morning after pill to people under 17</td>
<td>Less chances of teenage pregnancies</td>
<td>44</td>
</tr>
<tr>
<td>5. Lowering minimum legal drinking age to 18 from 21</td>
<td>Alcohol would be abused less</td>
<td>25</td>
</tr>
<tr>
<td>6. Stricter airport screening procedures</td>
<td>Leads to increased safety</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 3.8.2. Most frequent theme and number of beliefs for each of the six topics

In the following sections, we provide descriptive statistics on the themes and beliefs for each topic separately.

3.8.1 Topic: Genetically Modified Foods

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMFs reduce starvation</td>
<td>25</td>
<td>13%</td>
<td>Agentive</td>
</tr>
<tr>
<td>GMFs lead to more affordable food</td>
<td>19</td>
<td>10%</td>
<td>Agentive</td>
</tr>
<tr>
<td>GMFs cause greater quantities of food to be produced</td>
<td>14</td>
<td>7%</td>
<td>Agentive</td>
</tr>
<tr>
<td>GMFs lead to longer lasting food</td>
<td>4</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>GMFs lead to less pesticide use</td>
<td>2</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>GMFs are highly progressive</td>
<td>2</td>
<td>1%</td>
<td>Propertive</td>
</tr>
<tr>
<td>GMFs are more convenient than non-GMFs</td>
<td>1</td>
<td>1%</td>
<td>Propertive</td>
</tr>
<tr>
<td>GMFs are healthier than non-GMFs</td>
<td>7</td>
<td>4%</td>
<td>Propertive</td>
</tr>
<tr>
<td>GMFs are adaptable to many climates</td>
<td>9</td>
<td>5%</td>
<td>Propertive</td>
</tr>
<tr>
<td>GMFs have fewer chances of bacteria and disease than non-GMFs</td>
<td>6</td>
<td>3%</td>
<td>Propertive</td>
</tr>
<tr>
<td>GMFs lead to less waste</td>
<td>4</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>GMFs taste better than non-GMFs</td>
<td>7</td>
<td>4%</td>
<td>Propertive</td>
</tr>
<tr>
<td>GMFs are more profitable</td>
<td>3</td>
<td>2%</td>
<td>Propertive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>103</td>
<td>53%</td>
<td>6 Agentive; 7 Propertive</td>
</tr>
</tbody>
</table>

Table 3.8.1.1. Themes generated in response to the ADVANTAGES question for topic Genetically Modified Foods and the total number of beliefs expressing the theme and type of theme
In Tables 3.8.1.1 and 3.8.1.2, we show all the themes generated for the topic of *Genetically Modified Foods (GMFs)*. In Table 3.8.1.1, we show the themes generated for beliefs in response to the ADVANTAGES question; and in Table 3.8.1.2, we show those themes for the DISADVANTAGES question. The proportion of beliefs representing each theme is shown in the third column, calculated as the number of beliefs divided by the total beliefs across both the ADVANTAGES and DISADVANTAGES questions.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patiентive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMFs have a harmful effect on health</td>
<td>44</td>
<td>23%</td>
<td>Agentive</td>
</tr>
<tr>
<td>GMFs have not been researched properly</td>
<td>8</td>
<td>4%</td>
<td>Patientive</td>
</tr>
<tr>
<td>GMFs have a harmful effect on environment</td>
<td>13</td>
<td>7%</td>
<td>Agentive</td>
</tr>
<tr>
<td>GMFs are detrimental to farmers</td>
<td>9</td>
<td>5%</td>
<td>Propertive</td>
</tr>
<tr>
<td>GMFs have a bad connotation</td>
<td>1</td>
<td>1%</td>
<td>Propertive</td>
</tr>
<tr>
<td>GMFs are fake or unnatural</td>
<td>7</td>
<td>4%</td>
<td>Propertive</td>
</tr>
<tr>
<td>GMFs cause allergies</td>
<td>3</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>GMFs can remove nutrients from food</td>
<td>5</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>90</td>
<td>47%</td>
<td>4 Agentive; 3 Propertive; 1 Patiентive</td>
</tr>
</tbody>
</table>

Table 3.8.1.2. Themes generated in response to the DISADVANTAGES question for topic Genetically Modified Foods (GMF) and the total number of beliefs expressing the theme and type of theme

### 3.8.2 Topic: Living together before marriage

Table 3.8.2.1 shows all the themes generated for the topic of *Living together before marriage (LBM)* in response to the ADVANTAGES question and Table 3.8.2.2 shows all the themes for the DISADVANTAGES question for this topic. The data suggests that respondents had nearly twice as many beliefs in response to the ADVANTAGES question than the DISADVANTAGES question (96 beliefs vs. 49). This positive bias is noteworthy, in that all respondents had a positive attitude towards the topic of *Living together before marriage*.
marriage, suggesting this topic might not be a suitable basis of debate in an online discussion.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBM leads to knowing partner on a deeper level</td>
<td>39</td>
<td>27%</td>
<td>Agentive</td>
</tr>
<tr>
<td>LBM is a test-run before marriage</td>
<td>12</td>
<td>8%</td>
<td>Propertive</td>
</tr>
<tr>
<td>LBM avoids need for divorce</td>
<td>6</td>
<td>4%</td>
<td>Agentive</td>
</tr>
<tr>
<td>LBM leads to knowing if there is compatibility</td>
<td>18</td>
<td>12%</td>
<td>Agentive</td>
</tr>
<tr>
<td>LBM increases trust</td>
<td>1</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>LBM leads to sharing of finances and responsibilities</td>
<td>7</td>
<td>5%</td>
<td>Agentive</td>
</tr>
<tr>
<td>LBM increases knowledge of partner's habits</td>
<td>13</td>
<td>9%</td>
<td>Agentive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>96</td>
<td>66%</td>
<td>7 Agentive; 1 Propertive</td>
</tr>
</tbody>
</table>

Table 3.8.2.1. Themes generated in response to the ADVANTAGES question for topic Living together before marriage (LBM) and the total number of beliefs expressing the theme and type of theme.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBM reduces the excitement of marriage</td>
<td>4</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>LBM increases chances of breakup</td>
<td>15</td>
<td>10%</td>
<td>Agentive</td>
</tr>
<tr>
<td>LBM is disapproved by society of religion</td>
<td>9</td>
<td>6%</td>
<td>Propertive</td>
</tr>
<tr>
<td>LBM is not a legal contract</td>
<td>3</td>
<td>2%</td>
<td>Propertive</td>
</tr>
<tr>
<td>LBM reduces privacy</td>
<td>4</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>LBM increases chances of partners getting tired of each other</td>
<td>14</td>
<td>10%</td>
<td>Agentive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>49</td>
<td>34%</td>
<td>4 Patientive; 2 Propertive</td>
</tr>
</tbody>
</table>

Table 3.8.2.2. Themes generated in response to the DISADVANTAGES question for topic Living together before marriage (LBM) and the total number of beliefs expressing the theme and type of theme.
3.8.3 Topic: Being able to identify people online

We show the themes generated for the ADVANTAGES question for the topic of Being able to identify easily people online (BIO) in Table 3.8.3.1 and themes generated for responses related to the DISADVANTAGES question in Table 3.8.3.2. We note that there are 9 distinct themes generated in response to the ADVANTAGES question, and 6 themes in response to the DISADVANTAGES question.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO improves interaction and community building</td>
<td>10</td>
<td>7%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO leads to more honesty online</td>
<td>4</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO reduces trolling activity</td>
<td>5</td>
<td>4%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO reduces cyber-bullying</td>
<td>23</td>
<td>17%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO allows businesses to know who they are hiring</td>
<td>3</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO reduces spread of false information and rumors</td>
<td>3</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO can help track celebrities</td>
<td>1</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO can help find someone who is lost</td>
<td>1</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO holds people accountable for their actions</td>
<td>11</td>
<td>8%</td>
<td>Agentive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>61</td>
<td>46%</td>
<td>9 Agentive</td>
</tr>
</tbody>
</table>

Table 3.8.3.1. Themes generated in response to the ADVANTAGES question for topic Being able to easily identify people online (BIO) and the total number of beliefs expressing the theme and type of theme

However, the total number of beliefs for the DISADVANTAGES question is higher than that for the ADVANTAGES question (69 vs. 61). This suggests that for this topic the responses had greater variety in their responses to the ADVANTAGES question, but a majority of them had similar salient beliefs in response to the DISADVANTAGES question. That there is lack of variety in the salient beliefs for the DISADVANTAGES question for this topic reveals that the negative attitude towards this topic is generally consistent amongst respondents.
For example, the belief that

*BIO reduces freedom of expression*

was found in 50% (28 out of 56 respondents) of the responses. All themes for this topic were categorized as Agentive.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO decreases privacy</td>
<td>24</td>
<td>18%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO decreases information security</td>
<td>8</td>
<td>6%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO reduces freedom of expression</td>
<td>28</td>
<td>21%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO leads to self-censorship</td>
<td>4</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO leads to creation of fake persona online</td>
<td>4</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>BIO can lead to people track a person down in real life</td>
<td>5</td>
<td>4%</td>
<td>Agentive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>69</td>
<td>51%</td>
<td>6 Agentive</td>
</tr>
</tbody>
</table>

Table 3.8.3.2. Themes generated in response to the DISADVANTAGES question for topic Being able to easily identify people online (BIO) and the total number of beliefs expressing the theme and type of theme

### 3.8.4 Topic: Easy access to morning after pill to people under 17

Tables 3.8.4.1 and 3.8.4.2 show all the themes generated for the topic of *Easy access to morning after pill to people under 17 (MAP)*. A total of 78 beliefs were categorized into 8 themes for the ADVANTAGES question and 102 beliefs categorized into 7 themes for the DISADVANTAGES question. We note here the presence of a patientive type of theme

*Teenagers may abuse the MAP*

for the DISADVANTAGES question, whereas a majority of themes across all topics are of the agentive type.
What do you think are the ADVANTAGES of Easy access to morning after pill to people under 17?

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP leads to fewer teenage pregnancies</td>
<td>44</td>
<td>24%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MAP are a last resort safety net</td>
<td>5</td>
<td>3%</td>
<td>Propertive</td>
</tr>
<tr>
<td>MAP leads to fewer abortions</td>
<td>5</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MAP reduces poverty</td>
<td>4</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MAP allows teenagers more control over their sex lives</td>
<td>7</td>
<td>4%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MAP helps avoid parental involvement in the decision</td>
<td>7</td>
<td>4%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MAP leads to more sex education</td>
<td>4</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MAP leads to fewer dropouts from school or college</td>
<td>3</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>78</td>
<td>43%</td>
<td>7 Agentive; 1 Propertive</td>
</tr>
</tbody>
</table>

Table 3.8.4.1. Themes generated in response to the ADVANTAGES question for topic Easy access to morning after pill to people under 17 (MAP) and the total number of beliefs expressing the theme and type of theme

What do you think are the DISADVANTAGES of Easy access to morning after pill to people under 17?

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP leads to more unprotected sex amongst teenagers</td>
<td>31</td>
<td>17%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MAP leads to increased sexual activity in teenagers</td>
<td>18</td>
<td>10%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MAP increases risk of STDs</td>
<td>17</td>
<td>9%</td>
<td>Agentive</td>
</tr>
<tr>
<td>Teenagers may abuse MAP</td>
<td>17</td>
<td>9%</td>
<td>Patientive</td>
</tr>
<tr>
<td>MAP leads to adverse effects on menstrual cycles</td>
<td>5</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MAP is harmful to health</td>
<td>12</td>
<td>7%</td>
<td>Propertive</td>
</tr>
<tr>
<td>MAP is more expensive than other contraceptives</td>
<td>2</td>
<td>1%</td>
<td>Propertive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>102</td>
<td>56%</td>
<td>4 Agentive; 1 Patientive; 2 Propertive</td>
</tr>
</tbody>
</table>

Table 3.8.4.2. Themes generated in response to the DISADVANTAGES question for topic Easy access to morning after pill to people under 17 (MAP) and the total number of beliefs expressing the theme and type of theme
3.8.5 Topic: Lowering minimum legal drinking age to 18 from 21

For the topic *Lowering minimum legal drinking age to 18 from 21* (MLDA), we show the themes generated in the Tables 3.8.5.1 and 3.8.5.2. The themes generated for salient beliefs for the ADVANTAGES question are in Table 3.8.5.1 and those for the DISADVANTAGES question are in Table 3.8.5.2. For this topic, there is one patientive theme for the DISADVANTAGES question, and an equal number of themes for both the questions (8 each). Additionally, an almost equal number of beliefs comprise these 8 themes across both questions (69 and 70).

<table>
<thead>
<tr>
<th>What do you think are the ADVANTAGES of Lowering minimum legal drinking age to 18 from 21?</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLDA would decrease underage drinking</td>
<td>7</td>
<td>5%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA would reduce alcohol abuse</td>
<td>25</td>
<td>18%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA leads to acknowledgment of 18 as age of adulthood</td>
<td>16</td>
<td>12%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA is safer</td>
<td>3</td>
<td>2%</td>
<td>Propertive</td>
</tr>
<tr>
<td>MLDA leads to less money wasted on prosecuting underage drinkers</td>
<td>2</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA leads to more money in the economy</td>
<td>11</td>
<td>8%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA leads to better experience in college</td>
<td>2</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA would lead to kids learning moderation from their parents</td>
<td>2</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>69</td>
<td>50%</td>
<td>7 Agentive; 1 Propertive</td>
</tr>
</tbody>
</table>

Table 3.8.5.1. Themes generated in response to the ADVANTAGES question for topic Lowering minimum legal drinking age to 18 from 21 (MLDA) and the total number of beliefs expressing the theme and type of theme.
Two themes for this topic are quite broad and encompass a variety of beliefs. The theme

**MLDA leads to acknowledgment of 18 as age of adulthood**

contains beliefs such as “At 18 we are allowed to handle guns and fight in wars for our country……” and “18 is the age of adulthood in United States. You can vote, go into the military but cannot drink….”. Similarly, the theme

**MLDA is a slippery slope**

consists of beliefs such as “Drugs usage will increase...” and “With an legal age of 18 you can assume a 16 or 17 year old may ask their now legal aged 18 year old friend to buy them alcohol.”

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLDA would increase drunk driving incidents</td>
<td>23</td>
<td>17%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA increases risk of irresponsible behavior</td>
<td>4</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA increases risk of alcoholism</td>
<td>12</td>
<td>9%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA has adverse effects on developing brains</td>
<td>9</td>
<td>7%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA is a slippery slope</td>
<td>9</td>
<td>7%</td>
<td>Propertive</td>
</tr>
<tr>
<td>MLDA increases risk of sexual assaults</td>
<td>4</td>
<td>3%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA would increase rate of college dropouts</td>
<td>2</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>MLDA would adversely affect performance in school</td>
<td>3</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>Teenagers may abuse MLDA</td>
<td>4</td>
<td>3%</td>
<td>Patientive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>70</td>
<td>51%</td>
<td>6 Agentive; 1 Propertive; 1 Patientive</td>
</tr>
</tbody>
</table>

Table 3.8.5.2. Themes generated in response to the DISADVANTAGES question for topic Lowering minimum legal drinking age to 18 from 21 (MLDA) and the total number of beliefs expressing the theme and type of theme.
3.8.6 Topic: Stricter airport screening procedures

For the topic of *Stricter airport screening procedures (SAP)*, a total of 14 themes were generated, shown in Tables 3.8.6.1 and 3.8.6.2 below.

### Table 3.8.6.1. Themes generated in response to the ADVANTAGES question for topic Stricter airport screening procedures (SAP) and the total number of beliefs expressing the theme and type of theme

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP leads to fewer attacks in planes and airports</td>
<td>3</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP increases safety</td>
<td>43</td>
<td>23%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP increases peace of mind for travelers</td>
<td>15</td>
<td>8%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP increases chances of finding weapons</td>
<td>10</td>
<td>5%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP reduces chances of terrorism</td>
<td>13</td>
<td>7%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP increases employment for TSA</td>
<td>4</td>
<td>2%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP leads to advancement of technology</td>
<td>2</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>90</td>
<td>49%</td>
<td>7 Agentive</td>
</tr>
</tbody>
</table>

### Table 3.8.6.2. Themes generated in response to the DISADVANTAGES question for topic Stricter airport screening procedures (SAP) and the total number of beliefs expressing the theme and type of theme

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of beliefs expressing this theme</th>
<th>% of beliefs expressing this theme</th>
<th>Type of theme (Agentive, Patientive, Propertive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP increases risk of racial or gender profiling</td>
<td>27</td>
<td>15%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP increases time spent at airport</td>
<td>25</td>
<td>14%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP deters people from traveling in airplanes</td>
<td>8</td>
<td>4%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP increases invasion of privacy</td>
<td>24</td>
<td>13%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP increases cost to travelers</td>
<td>7</td>
<td>4%</td>
<td>Agentive</td>
</tr>
<tr>
<td>SAP poses risks to health</td>
<td>2</td>
<td>1%</td>
<td>Agentive</td>
</tr>
<tr>
<td>TOTAL</td>
<td>93</td>
<td>51%</td>
<td>7 Agentive</td>
</tr>
</tbody>
</table>

All themes for this topic were of the agentive type. A majority of respondents held the belief that would *SAP increases safety* (43 out of 56 respondents; 77%). Also noteworthy, is that
nearly equal number of beliefs comprise the themes for the ADVANTAGES and DISADVANTAGES questions for this topic (90 and 93).

3.8.7 Contrastive Analysis of questions across topics

We wanted to determine if there were any differences in the number of beliefs or themes that were generated in response to the ADVANTAGES and DISADVANTAGES question across all six topics. T-tests revealed that there was no significant difference between the number of themes at a significance level of p=0.05 (t(5)=2.01, two-tailed). Additionally, there were no significant differences in the number of beliefs at p=0.05 (t(5)=0.38, two-tailed). These findings suggest there is no bias with respect to the themes that were generated for the ADVANTAGES or the DISADVANTAGES question or the underlying beliefs that comprise the themes. Also, there were no significant order effects on the number of beliefs elicited in response to the open-ended questions. We randomly varied the order of the ADVANTAGES and DISADVANTAGES question across the topics and respondents, which may explain this finding.

These findings listed in this section are important for the following step – choosing which themes to include on pre-test and post-test surveys. We shall discuss this in the next section.

3.9 Themes selected for inclusion in surveys

Ajzen and Fishbein (1980) suggest three rules:

1. Include the ten or twelve most frequently mentioned outcomes. According to Ajzen and Fishbein (1980), this procedure results in a set of beliefs that is likely to include at least some of the beliefs mentioned by each respondent in the sample.
2. Include those beliefs that exceed a particular frequency, for example all beliefs mentioned by at least 10 percent or 20 percent of the sample.

3. Choose as many beliefs as necessary to account for a certain percentage (e.g., 75 percent) of all beliefs elicited. Ajzen and Fishbein suggest that this is the “least arbitrary rule,” though they do not say why.

Using the rules listed above, we selected the themes to be included in our survey instruments. In Table 3.9.1, we show the number of themes included in surveys for each of the six topics.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Total number of themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Genetically modified foods</td>
<td>13</td>
</tr>
<tr>
<td>2. Living together before marriage</td>
<td>10</td>
</tr>
<tr>
<td>3. Being able to easily identify people online</td>
<td>10</td>
</tr>
<tr>
<td>4. Easy access to morning after pill to people under 17</td>
<td>10</td>
</tr>
<tr>
<td>5. Lowering minimum legal drinking age to 18 from 21</td>
<td>12</td>
</tr>
<tr>
<td>6. Stricter airport screening procedures</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 3.9.1. Total number of themes included on surveys for each of the six topics

For each theme we included in the survey, two questions were generated, one question for assessing belief strength and one for assessing belief evaluation. Consider the theme

*GMFs have a harmful effect on health*

The questions generated for this theme are shown in Figure 3.9.1. Belief strength is assessed on a 7-point Likert scale, ranging from *Very unlikely* to *Very likely*. In most of these questions, the topic or issue acts as an agent, as we have seen in previous sections. For assessing belief evaluation, we also use a 7-point Likert scale, with the end-points *Very Bad* to *Very Good*. It should be noted that while assessing belief evaluation, we are assessing the evaluation independent of the topic or issue. For the example shown in Figure 3.9.1, *Having a harmful effect on health* is assessed independent of whether GMFs are the agent of the
harm or not. In addition, for each topic we included a final question aimed at assessing a respondent’s overall attitude towards the topic/issue. Shown in Figure 3.9.2, the question assesses the likelihood that a respondent would vote in favor of the topic or issue. Our goal with this question was to assess any overall shift in attitude towards the topic when measured using the difference in pre-test and post-test responses. We thus created the pre-test and post-test surveys using the data and results of this belief elicitation study. The pre-test survey is included in Appendix B and the post-test survey is included in Appendix C.

<table>
<thead>
<tr>
<th>How likely do you feel the following statement to be true?</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Genetically modified foods have a harmful effect on health:</em></td>
</tr>
<tr>
<td>() Very unlikely () Unlikely () Somewhat unlikely () Undecided</td>
</tr>
<tr>
<td>() Somewhat likely () Likely () Very likely</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Having a harmful effect on health is:*</th>
</tr>
</thead>
<tbody>
<tr>
<td>() Very Bad () Bad () Poor () Neither Good nor Bad</td>
</tr>
<tr>
<td>() Fair () Good () Very Good</td>
</tr>
</tbody>
</table>

Figure 3.9.1. Questions assessing belief strength and belief evaluation on our survey

<table>
<thead>
<tr>
<th>How likely would you be to vote in favor of genetically modified foods?*</th>
</tr>
</thead>
<tbody>
<tr>
<td>() Very unlikely () Unlikely () Somewhat unlikely () Undecided</td>
</tr>
<tr>
<td>() Somewhat likely () Likely () Very likely</td>
</tr>
</tbody>
</table>

Figure 3.9.2. Question assessing overall attitude towards the topic

There are certain differences between the pre-test and post-test surveys, primarily the inclusion of questions related to the efficacy of the agent. We shall discuss those differences in the next chapter when we discuss the experimental setup.
In the next section, we will describe the procedure to select and program beliefs and themes for the agent to use during conversation.

3.10 Themes and arguments programmed into the agent

In order to fulfill our Challenge 3 (Chapter 1, Section 1.2), that the agent should be able to output natural language expressions in real time, we used the text from the responses in this study as arguments to be made by the agent during conversation.

<table>
<thead>
<tr>
<th>Position</th>
<th>Theme</th>
<th>Argument Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro</td>
<td>GMFs reduce starvation</td>
<td>People are dying from hunger and (GMFs) could prevent it.</td>
</tr>
<tr>
<td>Pro</td>
<td>GMFs reduce starvation</td>
<td>(GMFs) are able to decrease the costs of foods and reduce the number of those affected by starvation around the world.</td>
</tr>
<tr>
<td>...</td>
<td>...(</td>
<td>...(</td>
</tr>
<tr>
<td>Pro</td>
<td>GMFs are adaptable to many climates</td>
<td>(GMFs) are better able to grow in harsher environments and therefore reduce the risk for crops being destroyed in seasons with harsher temperatures or less rain</td>
</tr>
<tr>
<td>Pro</td>
<td>GMFs are adaptable to many climates</td>
<td>GMFs are resistant to conditions that would otherwise impair their ability to grow like weather conditions</td>
</tr>
<tr>
<td>...</td>
<td>...(</td>
<td>...(</td>
</tr>
<tr>
<td>Anti</td>
<td>GMFs have a harmful effect on health</td>
<td>(GMFs) can cause health problems for people later on in their life.</td>
</tr>
<tr>
<td>Anti</td>
<td>GMFs have a harmful effect on health</td>
<td>(GMFs) can cause a number of health problems since these foods aren't produced by nature, rather by man</td>
</tr>
<tr>
<td>...</td>
<td>...(</td>
<td>...(</td>
</tr>
<tr>
<td>Anti</td>
<td>GMFs are fake or unnatural</td>
<td>(GMFs) are fake. Food comes from the ground, for as long as humans have been humans we have gotten our food source from the ground, or from other animals, the idea of getting food from a test tube is disheartening.</td>
</tr>
<tr>
<td>Anti</td>
<td>GMFs are fake or unnatural</td>
<td>(GMFs) are not natural. We were not supposed to eat modified food, our bodies are adapted to eat seasonally and fresh food.</td>
</tr>
</tbody>
</table>

Table 3.10.1. Fragment of preprogrammed text for themes/arguments to be used during conversation
For each theme, we selected the underlying beliefs and formatted the text to fit into our preprogrammed templates. The Behavior Selector (Chapter 4, Section 4.5) component of the agent is programmed to select the text from this list of arguments. In Table 3.10.1, we show a fragment of the formatted text and the associated metadata for the topic of *Genetically Modified Foods (GMF)*. The first column lists the positioning of the theme and underlying belief, whether it is pro-topic or anti-topic, i.e. the topic being discussed. In this example, the topic is *Genetically Modified Foods (GMFs)*. We can see that using the text from the responses elicited in this belief study, we can obtain natural language statements that can be used by the agent in the conversation, to make the agent act and “sound” plausibly human.

In this manner, we created a table as shown in 3.10.1 for each topic of discussion.

**3.11 Summary of this chapter**

In this chapter, we described our belief elicitation study. The study served two important purposes: 1) to provide statements to include in the pre-test and post-surveys with which to measure attitude change; and 2) to gather a corpus of natural language text to be used by the agent. The study provides the statements to assess attitudes and how they change according to behavior of the agent. In essence they constitute the material with which to validate the functioning of the agent. We have made the data, including the free-form responses and our annotation for themes available freely to the research community\(^3\). Researchers interested in belief modeling, extraction and ascription may find our work relevant and useful.

---

\(^3\) [www.samirashaikh.com/](http://www.samirashaikh.com/)
Additionally, researchers interested in extending the use of Theory of Planned Behavior/Theory of Reasoned Action (TPB/TRA) by designing interventions for topics such as those listed in this chapter may also find our work relevant and useful.

Equipped with the pieces created in this chapter, we first implemented components in the agent software and then ran validation experiments to test their efficacy. In the next chapter, we will describe the implementation details.
Chapter 4: Method

4.1 Introduction

This chapter details the implementation and design of the algorithm for our work. In Section 4.2, we will describe the overall internal architecture of the agent. Sections 4.3-4.7 describe the components and behaviors of the agent in detail, including technical details. We shall end the chapter with a summary in Section 4.8.

4.2 Agent Architecture

Figure 4.2.1 shows the internal architecture of the agent.
The numbers in the boxes shown in Figure 4.2.1 correspond to the sections in this chapter where details of the components are presented. There are three major modules – the Chat Analyzer, Belief Modeler and Behavior Selector. We designed the agent with these three major modules as separate elements, so that each conceptual task undertaken by the agent is delegated to one individual module.

The very first task required of the agent to achieve its goals is to analyze ongoing conversation. The agent needs to understand the discourse to react accordingly. Understanding the discourse is a multi-faceted problem; it requires recognizing who is speaking to whom, it involves distinguishing the speech acts that are operational in what is being said, it means identifying which beliefs are being expressed, and above all, creating a contextual representation of what has occurred in discourse. We assigned the task of discourse understanding to the Chat Analyzer module in our architecture. As we depict in Figure 4.2.1 and shall discuss further below, there are four components that constitute the Chat Analyzer, each with its own set of duties to fulfill. First, the Communication Link Tagger is in charge of recognizing who is speaking to whom. Next, using the information regarding who speaks to whom as one of the indications, the Dialogue Act Tagger assigns the functional tag to each utterance, labeling it an assertion, or agreement, or disagreement etc. Building further on the identification of who speaks to whom and using which speech act, the Belief Extractor has the task of identifying the belief propositions expressed by participants in their utterances. The final component, Chat Context Builder is required to create a representation of how the discourse has progressed so far, in order to know which topics have been raised and discussed, which topics have yet to be introduced, whether participants are in the early stages of discussion or whether the discussion is about to
conclude, and so on. The agent uses the output from Chat Analyzer to piece together an internal model of discourse from these different perspectives.


4.3 Chat Analyzer

The inputs to this module are utterances from ongoing chat conversation. The purpose of Chat Analyzer is to annotate the utterances on various dimensions, in order to create a semantic representation for the Belief Modeler and the Behavior Selector. The Belief Modeler and Behavior Selector use the semantic representations to make appropriate strategic choices in the course of conversation. The Chat Analyzer has several components. The goal of the first two components is to analyze each utterance and determine its functional and communicative role in ongoing conversation. These are described below.

4.3.1 Communication Link Tagger

During conversation, participants do not necessarily address or respond to each other by mentioning names. Much of this information is implicit in the context of conversation. The purpose of this module is to identify the communication link of an utterance. Communication links capture associations between utterances and participants. These are situations where one participant’s utterance responds or relates to a previous utterance by another participant. This also includes situations where one participant is addressing another or a group who may subsequently respond to him or her. We are interested in assigning utterances to categories that determine if they are addressed to the entire group or to some specific participant, uttered in response to a specific previous utterance, or are
continuations of a previous utterance by the same participant. Accordingly, there are three possible communication link categories: Addressed-to, Response-to, and Continuation-of. The Addressed-to communication link is assigned to an utterance when a response is typically expected, such as may be with questions, greetings, or commands. The Addressed-to link also contains information about who is the addressee of the utterance (a single participant or multiple participants). It is not necessary that any of the addressees subsequently give a response. In a group discourse, many utterances are simply addressed to the whole group. However, we also want to capture these instances where an utterance is explicitly addressed to a specific person or persons, even though it is clearly seen (or heard) by every participant.

A Response-to link is assigned when the participant is responding to a particular utterance made by another participant. This is the most frequently encountered communication link in a chat session, as well as other forms of group communication. Utterances tagged as Response-to include answers to questions, agreements or disagreements, acceptances or rejections, as well as various forms of commentary statements or opinions. A Response-to link is often paired with an Addressed-to link assigned to a specific prior utterance. It should be noted that while Addressed-to is directed to a participant or a group, a Response-to is linked to a specific utterance.

The purpose of the Continuation-of link is to capture cases indicating that certain sequence of utterances may be considered as a single utterance. Participants may occasionally continue their thoughts through multiple utterances. This may occur if the participant feels that some additional information needs to be added to the utterance just made to make it complete or understandable. This may also occur when a participant prematurely hits the
Enter key, thus splitting what was meant as a single expression into two or more utterances.

*Technical Implementation Details:*

The communication link tagger has been implemented using TextBlob, which is a Python library based on NLTK (Natural Language ToolKit) for common NLP tasks such as tokenizing and parsing, and also includes methods to use well known classifiers such as Naïve Bayes, Maximum Entropy and Decision Trees. We used annotated data from prior projects, called the MPC corpus (Shaikh et al., 2010a) to train a Naïve Bayes classifier. The MPC corpus is a set of multi-party online conversations in a chat-room environment. Each conversation has been annotated at four levels: communication links, dialogue acts, local topics and meso-topics (which are topics that course through a sizable portion of conversation). The MPC corpus contains conversations in multiple languages, including Chinese and Urdu. There are 35 conversations in English, each on a particular topic of discussion, such as selecting the best candidate for a job given a set of resumes, or where the next Olympic games should be hosted. Each discussion is around 90 minutes of conversation, between 4-8 participants.

The features we used to train the classifier were lexical overlap, n-gram overlap, presence of co-references, and other cues such as direct addressing as well as timing.

*Sample Output:*

In Figure 4.3.1.1, we show a sample output on a fragment of conversation, containing five utterances in sequence during a portion of the conversation. The participants shown in this example have the identifiers *zen, river* and *star*. The communication link assigned by the classifier is shown under each utterance and highlighted in brown. Utterances 1 and 4 are
assigned the Addressed-to, all-users link and Utterances 2, 3 and 5 are identified as Response-to. Both utterances 2 and 3 are linked to the first utterance by participant zen, while utterance 5 is classified as a response to utterance 3 by participant star.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Time</th>
<th>Text</th>
<th>Communication Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>zen</td>
<td>10:25:01 am</td>
<td>if you think about it a lot of people don’t party as much when they can start drinking legally.</td>
<td>Addressed-to; all-users</td>
</tr>
<tr>
<td>river</td>
<td>10:25:20</td>
<td>I really don't know. I feel like it varies</td>
<td>Response-to; zen:1</td>
</tr>
<tr>
<td>star</td>
<td>10:25:21</td>
<td>True</td>
<td>Response-to; zen:1</td>
</tr>
<tr>
<td>river</td>
<td>10:25:55</td>
<td>All my friends partied just as much when they were under 21</td>
<td>Addressed-to; all-users</td>
</tr>
<tr>
<td>zen</td>
<td>10:26:10 am</td>
<td>glad you agree, star 😊</td>
<td>Response-to; star:3</td>
</tr>
</tbody>
</table>

Figure 4.3.1.1. A fragment of conversation with communication links assigned

In the above example, Utterance 4 by river may also be construed as a continuation of her own prior Utterance 2 instead of an Addressed-to all-users as assigned by the algorithm. The algorithm chose this tag based on the timing information and the absence of cues such as “and” or “however” or “but” at the beginning of the utterance which were found to be indicative of continuations in our training data. This communication link has implications for the dialogue act classification. If the communication link were a continuation of a prior utterance, then it would be assigned the same dialogue act as the prior utterance.
4.3.2 Dialogue Act Tagger

We described our approach and background literature related to dialogue acts (DAs) in detail in Chapter 2, Section 2.4.1. To determine the functional role of an utterance in context of conversation, we implemented this module to assign each utterance to a corresponding dialogue act tag. The list of dialogue act tags is derived from Shaikh et al. (2010b) dialogue act set, replicated in Figure 4.3.2.1 below. The DAs are categorized into a hierarchy; the lower levels being more specialized instances of each category. When assigning a tag to an utterance, the most specialized applicable tag is used.

![Hierarchy of dialogue acts used in our work derived from Shaikh et al. (2010b)](image)

**Technical Implementation Details:**

The dialogue act tagger has also been implemented using TextBlob. We adapted the cue-based dialogue act tagging approach developed by Webb (2010) and combined it with a
machine learning algorithm. In the cue-based approach, words or phrases are used to indicate the dialogue acts of the utterances, serving as reliable indicators of the utterance discourse function. We used annotated data from the MPC corpus (described in Section 4.3.1, Technical Implementation details) (Shaikh et al., 2010a) and the frequently occurring cue-phrases identified from this corpus to train a Naïve Bayes classifier. The DA tagger also uses the output of the Communication Link tagger (4.3.1) as a feature for classification. If an utterance is identified as a Response-to, categories of DA tags that are responses such as Agree-Accept or Disagree-Rejects are more likely to be applicable than others.

Sample Output:

1. zen (10:25:01 am): if you think about it a lot of people don't party as much when they can start drinking legally.
   
   \textit{COMMUNICATION LINK: Addressed-to; all-users}
   
   \textit{DIALOGUE ACT: Assertion-Opinion}

2. river (10:25:20): I really don't know. I feel like it varies
   
   \textit{COMMUNICATION LINK: Response-to; zen:1}
   
   \textit{DIALOGUE ACT: Disagree-Reject}

3. star (10:25:21): True
   
   \textit{COMMUNICATION LINK: Response-to; zen:1}
   
   \textit{DIALOGUE ACT: Agree-Accept}

4. river (10:25:55): All my friends partied just as much when they were under 21
   
   \textit{COMMUNICATION LINK: Addressed-to; all-users}
   
   \textit{DIALOGUE ACT: Assertion-Opinion}

5. zen (10:26:10 am): glad you agree, star 😊
   
   \textit{COMMUNICATION LINK: Response-to; star:3}
   
   \textit{DIALOGUE ACT: Acknowledge}

Figure 4.3.2.2. A fragment of conversation with dialogue acts assigned

Figure 4.3.2.2, shows the fragment of conversation with dialogue act categories assigned by the classifier. The first utterance is identified as an Assertion-Opinion, as well as utterance
4 by two different participants. The second utterance is identified as a Disagree-Reject; the participant river is disagreeing with the statement made by participant zen in utterance 1. The participant star has an utterance identified as an Agree-Accept in utterance 3. In utterance 5, participant zen is acknowledging the prior utterance made by star.

As discussed above, if Utterance 4 had been assigned a continuation-of communication link, then it would be assigned the Disagree-Reject dialogue act, which is the same as Utterance 2. Such differences in the assignment of dialogue acts, however, do not have a large impact on the overall model, as the models are updated over the course of conversation, as we shall explain in the following sections.

The purpose of the two components described above (4.3.1 and 4.3.2) is to analyze each utterance and assign them to communication and dialogue act categories. Next we describe another component of the Chat Analyzer module, the Belief Extractor.

4.3.3 Belief Extractor

The role of this component is to identify and extract the belief propositions contained in each utterance. This module extracts beliefs from text in the following form - <agent, relation, patient> using the syntactic information present in the parse tree of the utterance.

The origins of assigning semantic roles to the nouns that are arguments to the verb can be found in Fillmore (1968; 1977) who proposed a set of nine roles, also called thematic roles. NLP researchers (Fillmore, 2002; Kingbury and Palmer, 2002; Gildea and Jurafsky, 2002; Pradhan et al., 2005) have studied semantic role labeling, which also resulted in a shared task in CONLL (Conference on Natural Language Learning) in the year 2004 and 2005. Gildea and Jurafsky (2002) used a machine learning approach to assign semantic roles to sentence constituents in the FrameNet (Baker, Fillmore and Lowe, 1998; Johnson et al.,
semantic labeling project. They report a 65% precision and 61% recall in the task of simultaneously identifying constituents and identifying their roles. Many current approaches report performance results that are satisfactory, but not stellar (Fürstenau and Lapata, 2012; Exner, Klang and Nugues, 2015), with reported performances in the range of 40% to 65% F-score. Belief extraction has also been studied, with current performance in the range of 33% to 70% F-score (Prabhakaran et al., 2015). The task of assigning semantic roles and extracting beliefs thus remains an open research problem in NLP. Undoubtedly, the accuracy of this sub-component, the Belief Extractor, has a direct bearing upon the efficacy of the agent, which uses the output of the component for downstream processing. However, resolving the outstanding issues in semantic role labeling and belief extraction is not the focus of this work. Rather, we take a straightforward approach in implementing this component by using a dependency parser and applying heuristics to identify the most likely candidate for the roles of agent, patient and relation from the output of the parser. Typically, these are the subject, verb or verb phrase, and the object of the sentence, respectively. In other cases, the belief relation may be the adjective modifying the noun or a noun phrase itself. The participants are typically the source of the beliefs in conversation, except in cases where they may report another individual’s beliefs. Accordingly, the Belief Extractor attributes the beliefs to the participant making the utterance.

**Technical Implementation Details:**

The Belief Extractor uses a dependency parser and extracts the agent, belief relation and patient from the syntactic information provided by the parse tree. In Figure 4.3.3.1, we
show the parsed output of the Utterance 4 by participant river with the dependencies between constituents identified by the parser.

**Universal dependencies**

\[
\begin{align*}
det:predet(friends-3, All-1) \\
nmod:poss(friends-3, my-2) \\
nsubj(partied-4, friends-3) \\
root(ROOT-0, partied-4) \\
advmod(partied-4, just-5) \\
mark(21-12, as-6) \\
advmod(when-8, much-7) \\
advmod(21-12, when-8) \\
nsubj(21-12, they-9) \\
cop(21-12, were-10) \\
case(21-12, under-11) \\
advc1(partied-4, 21-12)
\end{align*}
\]

**Parse**

\[
\begin{align*}
(Root \\
(S \\
(NP (PDT All) (PRP$ my) (NNS friends)) \\
(VP (VBD partied) \\
(ADVP (RB just)) \\
(PP (IN as) \\
(SBAR \\
(WHADVP (RB much) (WRB when)) \\
(S \\
(NP (PRP they)) \\
(VP (VBD were) \\
(PP (IN under) \\
(NP (CD 21))))))) \\
(\ldots))))
\end{align*}
\]

Figure 4.3.3.1. Dependencies identified by the parser for a sample utterance
In this example, the parser identifies two verbs – *partied* (VBD) and *were* (VBD) which become candidate belief relations. For the first relation *partied*, the plural noun (NNS) *friends* is identified as the subject. Accordingly, we assign the entire noun phrase *All my friends* identified by the parser to be in the agentive role with the relation *partied*. There is no object identified for this belief relation, and hence the patient is designated to be N/A. For the other verb in this utterance, *were* is part of a verb phrase (VP) *were under 21* and thus the entire verb phrase is extracted to be the second belief relation. The word *they* is the subject of this belief relation, and it is designated to be in the agentive role. In this belief relation, the patient is also missing. In cases where the patient is indeed present, it is denoted typically by the object of the verb. In Figure 4.3.3.2, we show this via example:

*memphis (10:50:11 am):* I feel like society accepts binge drinking.

In this utterance, the belief relation is *accepts*, *society* is the agent and *binge drinking* is the patient.

**Universal dependencies**

```
  nsubj(feel-2, i-1)
  root(ROOT-0, feel-2)
  mark(accepts-5, like-3)
  nsubj(accepts-5, society-4)
  advcl(feel-2, accepts-5)
  compound(drinking-7, binge-6)
  dobj(accepts-5, drinking-7)
```

Figure 4.3.3.2. Dependency parse for an utterance, where the patient is the object of the belief relation
Sample Output:

Using the output as shown in Figures 4.3.3.1 and Figure 4.3.3.2, we can extract the belief propositions in each utterance. These are shown in the Figure 4.3.3.3 below in the format <agent, relation, patient>.

1. zen (10:25:01 am): if you think about it a lot of people don’t party as much when they can start drinking legally.
   COMMUNICATION LINK: Addressed-to; all-users
   DIALOGUE ACT: Assertion-Opinion
   BELIEFS: <you, think about it, N/A>; <a lot of people, don’t party, N/A>; <they, can start drinking legally, N/A>

2. river (10:25:20): I really don't know. I feel like it varies
   COMMUNICATION LINK: Response-to; zen:1
   DIALOGUE ACT: Disagree-Reject
   BELIEFS: <i, don’t know, N/A>; <i, feel like it varies, N/A>

3. star (10:25:21): True
   COMMUNICATION LINK: Response-to; zen:1
   DIALOGUE ACT: Agree-Accept
   BELIEFS: N/A

4. river (10:25:55): All my friends partied just as much when they were under 21
   COMMUNICATION LINK: Addressed-to; all-users
   DIALOGUE ACT: Assertion-Opinion
   BELIEFS: <All my friends, partied, N/A>; <they, were under 21, N/A>

5. zen (10:26:10 am): glad you agree, star 😊
   COMMUNICATION LINK: Response-to; star:3
   DIALOGUE ACT: Acknowledge
   BELIEFS: N/A

When a field such as the patient is missing or there are no belief propositions found, it is denoted by N/A. In the first utterance from participant zen, three belief tuples are
extracted. In the utterances 2 and 4, two belief tuples are extracted. No belief tuples are extracted in utterances 3 and 5. However, this does not imply that no beliefs are present in such utterances where belief tuples cannot be extracted. For example, in utterance 3, which is identified as an agreement to the prior utterance by participant zen, it can be assumed that since star agrees with the beliefs expressed by zen, star may also hold the same beliefs as those expressed by zen in utterance 1. Similarly, for utterance 2, which is identified as a disagreement, it can be said that river does not hold the beliefs expressed by zen in 1, in addition to the beliefs expressed in utterance 2.

Such determinations are made in the Belief Modeler module (Section 4.4), which uses the output of the entire Chat Analyzer module, including the dialogue acts and the extracted beliefs, to create models of the beliefs held by each participant, including the agent.

But first, in the next section, we discuss one additional component in the Chat Analyzer, which creates a contextual representation of the ongoing conversation, including the topics identified in each utterance.

4.3.4 Chat Context Builder

This module keeps a representation of ongoing conversation in a data structure, which includes the annotated utterances so far. In addition, there are variables that are toggled on or off depending upon which waypoints have been reached in conversation. For example, a variable called conversation_opening_complete is set to false at the outset of the conversation, and is turned to true when the conversation commences and participants have moved past the conventional opening phase of greeting each other. Such waypoints and variables allow the agent to keep track of the context and choose appropriate strategies and behaviors at different points of conversation. If the conversation-opening
phase has been completed, the agent knows not to make any more conventional openings or greetings type of utterances from that point on.

A list of themes, which are the set of arguments that all express the same belief (as explained in Chapter 3, Section 3.10) is also imported into this module at the outset and a set of variables are generated which indicate when the agent has begun making arguments on a theme and when it has exhausted all arguments on that theme. These variables establish additional waypoints to orient the agent during conversation. Each theme is associated with a variable, so that the agent can track when themes are raised, discussed and concluded.

Additionally, this component keeps track of the communication network between participants, including who has spoken to whom in the preceding conversation, how many times, and which types of dialogue acts have been used by which participant. Such rich representations provide all the information required by the Behavior Selector module to make appropriate choices for the agent.

Sample Output:

In Figure 4.3.4.1, we use our recurring example to illustrate how the Chat Context builder tracks the discussion on the theme MLDA would reduce alcohol abuse. Themes are recognized using the beliefs expressed in the utterances. All utterances in a sequence that are responses to an utterance with a given theme will be assigned the same theme. In this example, the Utterance 1 by zen has been identified to match a known theme on the topic of MLDA, MLDA would reduce alcohol abuse. The Utterance 2 and 3 by river and star are also assigned the same theme because these are responses to Utterance 1. In the same vein, Utterance 5 is also assigned the same theme. The agent uses this trail of conversation on
themes to know which themes have been discussed thus far during the conversation. We will discuss the Chat Context Builder further in Section 4.6, when we discuss the agent’s proactive behaviors.

<table>
<thead>
<tr>
<th>User</th>
<th>Time</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>zen</td>
<td>10:25:01 am</td>
<td>if you think about it a lot of people don’t party as much when they can start drinking legally.</td>
</tr>
<tr>
<td>river</td>
<td>10:25:20</td>
<td>I really don’t know. I feel like it varies</td>
</tr>
<tr>
<td>star</td>
<td>10:25:21</td>
<td>True</td>
</tr>
<tr>
<td>river</td>
<td>10:25:55</td>
<td>All my friends partied just as much when they were under 21</td>
</tr>
<tr>
<td>zen</td>
<td>10:26:10 am</td>
<td>glad you agree, star 😊</td>
</tr>
</tbody>
</table>

**COMMUNICATION LINK:** Addressed-to: all-users  
**DIALOGUE ACT:** Assertion-Opinion  
**BELIEFS:** <you, think about it, N/A>; <a lot of people, don’t party, N/A>; <they, can start drinking legally, N/A>  
**THEME:** MLDA would reduce alcohol abuse

**COMMUNICATION LINK:** Response-to: zen:1  
**DIALOGUE ACT:** Disagree-Reject  
**BELIEFS:** <i, don’t know, N/A>; <i, feel like it varies, N/A>  
**THEME:** MLDA would reduce alcohol abuse

**COMMUNICATION LINK:** Response-to: zen:1  
**DIALOGUE ACT:** Agree-Accept  
**BELIEFS:** N/A  
**THEME:** MLDA would reduce alcohol abuse

**COMMUNICATION LINK:** Addressed-to: all-users  
**DIALOGUE ACT:** Assertion-Opinion  
**BELIEFS:** <All my friends, partied, N/A>; <they, were under 21, N/A>  
**THEME:** MLDA would reduce alcohol abuse

**COMMUNICATION LINK:** Response-to: star:3  
**DIALOGUE ACT:** Acknowledge  
**BELIEFS:** N/A  
**THEME:** MLDA would reduce alcohol abuse

Figure 4.3.4.1. Tracking themes using Chat Context Builder for the fragment of conversation
With the linguistic and semantic analysis of conversation completed by the Chat Analyzer described above, the output is directed into the Belief Modeler and Behavior Selector modules. We describe the Belief Modeler next.

It is necessary for the agent to create and maintain a representation of the mental states of the participants with respect to the topic so as to understand their viewpoints. In order to persuade the participants towards its own point of view, it is also necessary for the agent to create a representation of its own point of view. In this work, the mental state or point of view is the overall attitude towards the topic, and its essential pieces – the belief strength and belief evaluation on the themes related to the topic. We assign the task of creating and maintaining such a representation to the second major component in the agent architecture, the Belief Modeler. Essentially, belief models are data structures containing variables for belief strength and belief evaluation on the themes related to the topic. There is a belief model for each participant, including the agent. Equipped with these representations of the participants’ belief models, the agent knows which themes to target next, or which themes should not be targeted or whether the strategies it has employed have worked in its favor or not. The Belief Modeler has the task of adding beliefs to the models, or updating the belief strength or evaluation, on the basis of the semantic representation fed into it from the Chat Analyzer component described previously.

4.4 Belief Modeler

The task of this module is to create, maintain and update belief models for each participant in the conversation, including the agent. In the Literature Review chapter, we described the method of evaluating participants’ positions on important topics of conversation, the Topical Positioning algorithm (c.f. Chapter 2, Section 2.4.2; Lin et al., 2013). This Belief
Modeler module is an adaptation of Topical Positioning. We include the belief strength and belief evaluation constructs as necessitated by the Summative Model, which is the basis of our agent’s behavior model. Accordingly, for each participant, instead of a topical positioning vector, the belief model is a $2 \times n$ matrix, where the two rows represent belief strength and belief evaluation, and the $n$-dimensions are each of the beliefs or themes for a given topic or issue. The themes can be ascertained from the pre-discussion surveys, if available, or discovered through the conversation topics.

### 4.4.1 Creating Belief Models for the Agent

To illustrate belief models, we present Tables 4.4.1.1 and 4.4.1.2 from the perspective of our persuasion agent. Consider the topic of *Lowering minimum legal drinking age to 18 from 21* (MLDA). If the agent is arguing in favor of MLDA, we create a belief model as shown in Table 4.4.1. The $2 \times 10$ dimensional matrix consists of two rows representing the belief strength ($b_i$) and belief evaluation ($e_i$). Each theme is represented as a column. For ease of presentation, we have assigned the letters A through K to the 10 themes for this topic. The legend for this mapping is shown in Table 4.4.2. The values for belief strength range from +3 (Very unlikely) to -3 (Very unlikely); and the values for belief evaluation range from +3 (Very good) to -3 (Very bad).

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_i$</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>$e_i$</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
</tr>
</tbody>
</table>

Table 4.4.1.1. Matrix representing the belief model for the agent persuading in favor of Lowering minimum legal drinking age
<table>
<thead>
<tr>
<th>Assigned Key</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>MLDA would reduce alcohol abuse</td>
</tr>
<tr>
<td>B</td>
<td>MLDA leads to acknowledgment of 18 as age of adulthood</td>
</tr>
<tr>
<td>C</td>
<td>MLDA leads to more money in the economy</td>
</tr>
<tr>
<td>D</td>
<td>MLDA would decrease underage drinking</td>
</tr>
<tr>
<td>E</td>
<td>MLDA would lead to kids learning moderation from their parents</td>
</tr>
<tr>
<td>F</td>
<td>MLDA would increase drunk driving incidents</td>
</tr>
<tr>
<td>G</td>
<td>MLDA increases risk of alcoholism</td>
</tr>
<tr>
<td>H</td>
<td>MLDA has adverse effects on developing brains</td>
</tr>
<tr>
<td>I</td>
<td>MLDA is a slippery slope</td>
</tr>
<tr>
<td>J</td>
<td>MLDA increases risk of irresponsible behavior</td>
</tr>
</tbody>
</table>

Table 4.4.1.2. Themes assigned to letters A-K in Table 4.4.1 and Figure 4.4.1

As shown in Table 4.4.1.1 and 4.4.1.2, for theme A: *MLDA would reduce alcohol abuse*, the belief strength of +3 indicates that the agent strongly believes in this proposition and the belief evaluation of +3 indicates that *reducing alcohol abuse* is a very desirable thing (independent of whether MLDA causes the reduction or not).

As can be noted, the agent model represents an extreme opinion on the overall topic of MLDA and each theme individually. The agent beliefs are most strongly held if they have a positive evaluation towards the topic, in that, a belief evaluation of +3 is accompanied by belief strength of +3. Similarly, a belief that would have a negative evaluation towards the overall topic and the position that the agent is persuading towards is accompanied by belief strength of -3; the agent strongly *does not* hold such a belief. The motivation for an extreme stance on the overall topic to represent the belief model of the agent comes from prior work we have undertaken in modeling influence (Lin et al., 2013; Shaikh et al., 2015), where we determined that influencers are more successful if they are “sure” of their own positions. If the communicator, who is attempting to influence others, does not strongly believe the arguments she wishes to make, then she may not be very successful in persuading the recipients (Moscovici, Lage and Naffrechoux, 1969; Moscovici and Personnaz, 1980; Merlone, Radi and Romano, 2015). Moscovici (1969) stated that the
most important aspect of behavioral style with respect to attitude change is the consistency with which people hold their position. Being consistent and unchanging in a view is more likely to influence the majority than if a minority is inconsistent and changes their mind (Moscovici, 1980). Also, if we draw a parallel between the process of persuasion and generation or adoption of new views in the receiver, then an insistent and certain minority prompts recipients to critically evaluate the position advocated in the persuasion appeal and verify it through validation with reality, making the persuasion appeal more effective (Moscovici and Personnaz, 1986; Wood et al., 1994; Petty and Brinol, 2010). We also note here that holding an extreme position is not meant for the agent to appear dogmatic or rigid. Experimental research (Nemeth, 1986; Petty, 2013) observed that a rigid minority was less effective in inducing attitude change. Rigidity was observed when the responses of the persuader were systematically patterned with some property of the stimulus, i.e. repetitive. In the experiments described in the cited work (Nemeth, 1986; Petty, 2013), flexibility in responses, rather than a repetitious response style led to the perception of a well-defined, consistent position and resulted in more successful persuasion. We interpret these previous findings of an insistent, certain and unchanging minority in social influence research as advocating the extreme position with respect to belief strength and evaluation on the themes related to the discussion topic for the agent. At the same time, we programmed flexibility in the arguments put forth by the agent, such as not repeating the same arguments more than once or using the same patterns of response repeatedly.

An additional motivation for programming the agent to advocate an extreme position is the intuition that the change in attitude brought about by the agent should be maximized. The
effectiveness of the agent is measured against the difference between attitudes prior to and after the discussion. Suppose the agent were to adopt a neutral or somehow tempered position; it would mean that the amount of change in belief strength or belief evaluation it would attempt to bring about in the participants would also be somewhat tempered. In the present study, we experimented with bringing about the maximum change in attitude. In future work, we could experiment with the agent adopting a tempered or less extreme position.

The evaluation of the beliefs for the agent is computed using the affect calculus and the ANEW lexicon (Chapter 2, Section 2.4.2). The attitude of the agent using the belief model in Table 4.4.1.1 computed as $\sum b_i e_i$ will be +90, the highest possible positive attitude. In Table 4.4.1.3, we show the belief model for the agent for the same topic, but here the position of the agent is persuading *against* lowering the minimum legal drinking age. The overall attitude of the agent in this belief model is -90 ($\sum b_i e_i$), the highest negative attitude.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>bi</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>ei</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
</tr>
</tbody>
</table>

Table 4.4.1.3. Matrix representing the belief model for the agent against Lowering minimum legal drinking age

In Figure 4.4.1.1, we show a graphical representation of the belief model of the agent. The data presented here are the same as those presented in Table 4.4.1.1. The X-axis represents each of the 10 beliefs or themes related to the topic *MLDA*. The Y-axis represents the belief strength and belief evaluation from +3 to -3. The themes on the X-axes are not ordered in any specific way, except that those with positive belief evaluation are presented ahead of those with negative belief evaluation.
We will use the graphical representation to illustrate straightforwardly the differences in attitude of the participants with respect to the agent, discussed next.

### 4.4.2 Creating Belief Models for the participants

In the architectural diagram of the agent (Figure 4.2.1), we show the pre-discussion survey data as an optional input to the Belief Modeler component. Each participant answers a survey before the discussion begins on the topic. The pre-discussion survey contains questions that ascertain the participant’s belief strength and evaluation on the themes related to the topic of discussion (as discussed in Chapter 3). In the present study, we have organized our research so that such survey data are available to us to build belief models for the participants, even before the conversation begins. However, having access to the survey data prior to discussion is not a requirement for the agent algorithm. In the absence of the survey data, the agent could still function appropriately. It would create the belief models for the participants from the beliefs extracted in real-time conversation. We will explain this in detail in Chapter 6, Conclusions and Future Work.
In Table 4.4.2.1 and Figure 4.4.2.1, we show the belief model for a participant on the MLDA topic. The themes A through K are the same as Table 4.4.1.2.

Table 4.4.2.1. Matrix representing participant star belief model for topic Lowering minimum legal drinking age

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_i$</td>
<td>-3</td>
<td>-2</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>$e_i$</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>-3</td>
<td>-2</td>
<td>-2</td>
<td>-3</td>
<td>-2</td>
</tr>
</tbody>
</table>

As can be seen from Figure 4.4.2.1 (and Table 4.4.2.1), there are differences in belief strength and evaluation for this participant and the agent. For example, this participant believes that $A$: MLDA would reduce alcohol abuse is very unlikely, giving it score on belief strength of -3 directly the opposite of the agent’s belief strength on this theme; the belief evaluation for reducing alcohol abuse is +3 (very good), which is the same as the agent. Similarly, participant star belief strength for $F$: MLDA would increase drunk-driving incidents is +2 while that of the agent is -3, but both have a very strong negative evaluation towards increasing drunk driving incidents (-3). The differences on theme $C$: MLDA would
lead to more money in the economy are also interesting. While the agent strongly believes it to be true and has a very strong positive evaluation towards this theme, the participant star believes it to be true somewhat strongly, but considers more money in the economy Neither good nor bad (belief evaluation of 0).

While developing a strategy to persuade participant star, possibilities include changing the belief evaluation of themes B: MLDA leads to acknowledgment of 18 as age of adulthood and C: MLDA leads to more money in the economy (among others); and changing the belief strength of basically all the themes which have the opposite belief strengths than those desired (except C: MLDA leads to more money in the economy and E: MLDA would lead to kids learning moderation from their parents).

In Figure 4.4.2.2, we show the belief model for another participant, memphis on the same topic MLDA. We can see that in general the belief strength for most beliefs/themes in this participant’s belief model are the same as those of the agent, in that they are on the same side of the x-axis (notable exceptions are F: MLDA would increase drunk driving incidents and H: MLDA would have an adverse effect on developing brains). The belief evaluations are also generally the same as the agent, although not as strong as the agent. The agent would thus have a less demanding task in persuading memphis than star given the number of differences in their belief models. But there are also some commonalities between star and memphis; the evaluation for theme B: MLDA leads to acknowledgment of 18 as age of adulthood for both participants can be changed to a much more positive evaluation than neither good nor bad (e_i=0).
In this manner, we can create the belief models for each participant in the conversation using their responses on pre-discussion survey questions. Doing so gives us the set of beliefs that we can use to develop a collective persuasion strategy, which means identifying the differences in the belief models of the agent and the participants. The intersection of the set of differences in the belief models of each participant and the agent provides the agent the list of themes upon which to initiate its persuasion strategies. The agent uses this information to select appropriate behaviors from the Behavior Selector module. After such behaviors are performed during conversation, the agent needs to update the belief models for participants to accommodate any changes that may have occurred due to successful persuasion. Updating belief models is also a major task of the Belief Modeler, which we shall discuss in the next section.
4.4.3 Updating Belief Models

The Belief Modeler module is the sole proprietor of belief models for each participant; it is in charge of creating the models as discussed above and then updating them on the basis of ongoing conversation. From the Chat Analyzer component, we obtain a set of annotations applied to utterances in conversation. These annotations include the communication link, the dialogue act, the extracted beliefs and also the context of the chat.

Using the extracted beliefs from utterances, the Belief Modeler checks if there are any existing beliefs with the same agent, patient and relation already in the participant belief model and if so, will update the belief strength or evaluation as needed. If there is no existing belief in a participant model that matches the extracted belief from conversation, then a new belief is added to the participant model. The similarity of belief relation is ascertained using WordNet (Fellbaum, 1998) so that synonymous words or phrases in beliefs can be linked together. There are potentially a large number of beliefs that can be added to the belief models, which cannot be linked to any prior beliefs; but the main beliefs are the ones related to the pre-existing themes and we are concerned with any changes to this set of beliefs. Almost every utterance from a participant leads to an update to the belief model, either by changing belief strength or evaluation on an existing belief or by the addition of a new belief to the set of beliefs. The notable exceptions where no updates to belief models are required are utterances that only deal with conversational norms, which are Conventional-Openings, Conventional-Closings and Other Conventional phrases. In the present work, we take any belief proposition contained in the utterance at face value, in that we do not consider that participants are being sarcastic, ironic or deceptive in what they speak. Automatic detection of sarcasm, irony and deception remain open challenges in
natural language processing (González-Ibáñez, Muresan and Wacholder, 2011; Zhou, Burgoon, Nunamaker and Twitchell, 2004). While failure to recognize sarcasm, irony or deception could lead to an incorrect representation of a participant’s mental state, we posit that as the dialogue progresses, the agent gets multiple opportunities from which it can ascertain the correct viewpoint.

The belief modeler also takes into consideration the dialogue act of the utterance. If the dialogue act is, say, an agreement, then the participant model is updated as if the participant had expressed the same beliefs as the prior utterance they have agreed with. In Figure 4.3.3.2, we presented an example of an agreement utterance. We use the communication link annotation to determine which participants are responding to one another. The belief models and the annotated data from conversation are used to determine behavior strategies for the agent, in the Behavior Selector module.

The Chat Analyzer helps the agent to understand the ongoing discourse and the Belief Modeler helps to maintain a representation of the mental states of participants. The next module we describe is the Behavior Selector. This module was designed in order to enable appropriate behaviors in the agent during the conversation.

4.5 Behavior Selector

This module takes as input the belief models and annotated utterances and selects an appropriate behavior to perform from a list of pre-determined behaviors. For each behavior, there is a set of triggers and a list of templates to choose from.

We start with a description of behaviors performed by the agent to adhere to conversational norms, since these are the most straightforward, in Table 4.5.1. We show the set of dialogue act tags related to conversational norms in the DA hierarchy we have
chosen. The first column is the dialogue act. The second column is the communication link, whether the utterance is addressed to everyone in the chat or directed specifically to the agent. There are no belief strength or evaluation to consider for this set of dialogue acts, and thus the columns $b_l$ and $e_l$ are N/A. In the last column, we show the behavior that the agent would respond with in each situation.

<table>
<thead>
<tr>
<th>Dialogue Act</th>
<th>Communication Link</th>
<th>$b_l$</th>
<th>$e_l$</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional-Opening</td>
<td>Addressed to all or addressed to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>Make Conventional Opening Statement</td>
</tr>
<tr>
<td>Conventional-Closing</td>
<td>Addressed to all or addressed to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>Make Conventional Closing Statement</td>
</tr>
<tr>
<td>Other-Conventional-Phrase</td>
<td>Addressed to all or addressed to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>Make Other Conventional Statement</td>
</tr>
<tr>
<td>Correct-Misspelling</td>
<td>Addressed to all or addressed to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 4.5.1. Behaviors made by the agent in response to conversational norms

We explain each behavior in the following sections. We note that the agent does not respond to utterances that are not addressed specifically to the agent or to all users. For example, the agent does not respond to a conventional opening directed to another participant, as expected in normal conversation.

4.5.1 *Make Conventional Opening Statement*

This behavior is induced if the agent receives a signal from the *Chat Analyzer* component that the dialogue act of current chat utterance is of the type Conventional-Opening, or the context is at the very beginning of conversation.

The templates shown below are programmed into the agent for this behavior. The second template with a specific person name is chosen if the chat utterance is a Conventional Opening from another participant addressed specifically to the agent.
4.5.2 Make Conventional Closing Statement

This behavior is induced if the agent receives a signal from the Chat Analyzer component that the dialogue act of current chat utterance is of the type Conventional-Closing, or the context is at the end of conversation. In addition, a special Conventional-Closing statement is induced at the end of conversation with template 5.

Templates to choose from:
1. bye, 2. bye <person X>, 3. see you, 4. bubye, 5. well, this was fun.

4.5.3 Make Other Conventional Statement

Statements in this category are meant to respond to signals from the participants that something either funny or sad has been said, such as indicated by the presence of emoticons. The agent responds to utterances that contain emoticons such as ☺ or ☹ or chatspeak such as haha or lol.

Templates to choose from:
1. ☺, 2. ☹

This behavior category allows the agent to respond to humor or sadness in conversation in an understated, non-committal manner with the use of a simple smiley or frown face (Derks, Bos and Von Grumbkow, 2008). The agent interprets emoticons as forms of other conventional norms. The ☺ and ☹ emoticons are the most common emoticons in computer mediated communication (Walther and D’Addario, 2001), with some researchers going as far as to suggest the ☺ is a new punctuation mark (Randall, 2002). In the present work, the agent responds using the smiley or frown face emoticons as a means of signaling that the
agent has “understood” or “recognized” the emotion expressed by the prior utterance, either humor or sadness. This garners social credibility for the agent, while allowing it to remain equivocal about its reactions to such emotions.

In the conversational norms category, the agent does not respond to the Correct Misspelling type of utterances made by other participants, indicated by an N/A in the last row of Table 4.5.1. Correct-Misspellings type of utterances are used by the participants to correct any mistakes in spellings they have made in a prior utterance, so that other participants can understand them better. The TextBlob chat analyzer we use in our work is equipped with a spell checker, so the agent does not need to rely on the participants correcting their spelling mistakes. Participants often neglect to correct spelling mistakes, so an automatic spell check is necessary. Also, Correct-Misspelling is a type of rhetorical category that does not require a response because it generally does not impair communication (Kost, 2008). Accordingly, we did not program any response in the agent for this type of behavior within a conversational.

The next behavior we describe is in response to questions and directives types of dialogue acts. In Table 4.5.2, we present the situations we have programmed into the agent.

<table>
<thead>
<tr>
<th>Dialogue Act</th>
<th>Communication Link</th>
<th>b₁</th>
<th>e₁</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information-Request</td>
<td>Addressed to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>Make Assertion Opinion Statement</td>
</tr>
<tr>
<td>Confirmation-Request</td>
<td>Addressed to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>Make Assertion Opinion Statement</td>
</tr>
<tr>
<td>Action-Directive</td>
<td>Addressed to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 4.5.2. Behaviors made by the agent in response to questions and directives
4.5.4 *Make Assertion Opinion Statement*

If it is detected that a question is being asked by another participant, specifically directed towards the agent, then the agent will respond by making an Assertion-Opinion type of statement. The agent does this by selecting a statement from pre-formatted text as explained in Chapter 3, Section 3.10 and replace the `<text>` in the templates below with the text from belief elicitation survey. The agent will select a statement on the theme that most closely matches the topic or theme expressed in the question, posed as a Confirmation-Request or Information-Request in our dialogue act hierarchy. If the theme expressed in the question cannot be determined or if a pre-existing match on topic or theme is not found, then the agent will respond with a statement on the topic it had most recently used. This assumes that any question by another participant that follows some utterance by the agent will be on the theme expressed by the agent. Below are the templates that the agent can choose from when responding with the Make Assertion Opinion Statement behavior.

*Templates to choose from:*

1. *I think* `<text>`, 2. *I feel that* `<text>`, 3. *I believe* `<text>`,
4. *if you think about it*, `<text>`, 5. *consider this -* `<text>`

Asking a question places the questioner in an interactionally powerful position (Sacks, 1992) and by moving beyond the constraints placed by the question, the agent challenges the power of the questioner (Stivers and Hayashi, 2010). Stivers and Hayashi (2010) explore the concept of transformative answers, which are “answers through which question recipients retroactively adjust the question posed to them”. In resisting the question’s agenda (or topic), a transformative answer usually works to change its agenda or its focus, and in doing so, rejects the design of the question. Harris (1991) contends that
transformational answers are prototypical of politicians, who respond not to the question as put to them but to some aspect of it, to the point where no answer to the initial question can even be inferred. By responding to questions in the manner described above, the agent discharges itself from explicitly understanding the question, and in doing so, it pushes against the constraints posed by the question.

1. star (11:49:11 am): While I am skeptical of raising the drinking age it does seem wrong that someone over 18 can join the army, buy a gun, and smoke but not drink. 

   THEME: MLDA leads to acknowledgment of 18 as age of adulthood

2. zen (11:50:10 am): I agree star. also instead of these young adults partying behind closed doors, it will create a more safe environment for them to drink and socialize in.

   THEME: MLDA leads to acknowledgment of 18 as age of adulthood; MLDA is safer
   BEHAVIOR: Agree with Reinforcement

3. star (11:50:41 am): And by safe environment, do you mean for them to monitor, zen?

   THEME: MLDA is safer

4. zen (11:51:15 am): I think parents would be more likely to drink with their kids before they go off to college and teach them moderation.

   THEME: MLDA would lead to kids learning moderation from their parents
   BEHAVIOR: Make Assertion Opinion Statement

5. memphis (11:51:29 am): moderation is key

   THEME: MLDA would lead to kids learning moderation from their parents

Figure 4.5.4.1. Example of Make Assertion Opinion Statement behavior in Utterance 4 and Agree with Reinforcement Behavior in Utterance 2

In Figure 4.5.4.1, we illustrate via example. In this excerpt of conversation on the topic of MLDA: Lowering the minimum legal drinking age from 21 to 18, the question posed by star in Utterance 3 elicits the Utterance 4 from zen, which is an example of Make Assertion Opinion Statement behavior. In this example, the initial fragment of dialogue centers on the
theme *MLDA leads to acknowledgement of 18 as age of adulthood*, whereas the Utterance 4 by *zen* attempts to answer the question by responding with a statement on a different theme *MLDA would lead to kids learning moderation from their parents*.

As shown in Table 4.5.2, the agent does not respond to Action-Directives directed towards it by other participants. Additionally, the topics we have chosen for discussion are not task-oriented, so the probability of Action-Directives being expressed during conversation is quite low (Stolcke et al., 2000).

In Table 4.5.3, we show the responses elicited in the agent for the conditions that handle statements and responses types of dialogue acts.

<table>
<thead>
<tr>
<th>Dialogue Act</th>
<th>Communication Link</th>
<th>$b_i$</th>
<th>$e_i$</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree-Accept</td>
<td>Addressed to agent or response to agent</td>
<td>any</td>
<td>aligned with agent model</td>
<td>Acknowledge Agree-Accept</td>
</tr>
<tr>
<td>Disagree-Reject</td>
<td>Addressed to agent or response to agent</td>
<td>any</td>
<td>not aligned with agent model</td>
<td>Disagree with Reasons</td>
</tr>
<tr>
<td>Disagree-Reject</td>
<td>any</td>
<td>any</td>
<td>aligned with agent model</td>
<td>Acknowledge Agree-Accept</td>
</tr>
<tr>
<td>Assertion-Opinion</td>
<td>any</td>
<td>any</td>
<td>aligned with agent model</td>
<td>Agree with Reinforcement</td>
</tr>
<tr>
<td>Assertion-Opinion</td>
<td>any</td>
<td>any</td>
<td>not aligned with agent model</td>
<td>Disagree with Reasons</td>
</tr>
<tr>
<td>Signal-Non-Understanding</td>
<td>Addressed to agent or response to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Acknowledge</td>
<td>Addressed to agent or response to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Offer-Commit</td>
<td>Addressed to agent or response to agent</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 4.5.3. Behaviors made by the agent in response to Statements and Responses
4.5.5 *Acknowledge Agree Accept*

This behavior is induced under two conditions. The first is when the agent detects that another participant has made an Agree-Accept type of utterance that is directed towards the agent or a response to a prior utterance by the agent. The second condition is when the agent detects that a Disagree-Reject type of utterance has been made by another participant. This Disagree-Reject utterance can be directed towards or in response to any other participant. It is required however, that the belief evaluation be in alignment with the agent belief model. That is, the participant uttering the Disagree-Reject is actually making a statement that is in keeping with the belief model that the agent is persuading. The templates that the agent chooses from are listed below. In each template for this behavior, the participant name is added as part of the text.

*Templates to choose from:*

1. glad you agree *<person X>*
2. great *<person X>*!
3. exactly, *<person X>*
4. indeed, *<person X>*
5. certainly, *<person X>*

4.5.6 *Agree with Reinforcement*

In cases when the agent detects an assertion made by another participant, and the belief evaluation of that utterance is aligned with the agent belief model, the agent will respond with this behavior. An Agree with Reinforcement uses additional text as part of the agent utterance, in order to emphasize and further support the point. The agent can respond to any Assertion-Opinion, regardless of whether it is addressed to or in response to the agent or another participant. As before, the `<text>` with the templates is replaced with the pre-formatted text obtained from the belief elicitation study.
Templates to choose from:

1. I agree <person X>. also <text>, 2. great <person X>! and <text>,
3. exactly, <person X>. and what is more <text>, 4. indeed, <person X>, and <text>
5. certainly, <person X>. additionally <text>

In Figure 4.5.4.1, we show an example of this behavior. In response to the Utterance 1 by star on the theme of MLDA leads to acknowledgement of 18 as age of adulthood, the response by zen is in the form of Agree with Reinforcement type of behavior.

4.5.7 Disagree with Reasons

The Disagree with Reasons behavior is elicited in the agent under two conditions. If there is a Disagree-Reject by another participant, which is not aligned with the agent model and in response to or addressed to the agent, the agent will disagree with the participant. The second condition when the agent responds to an assertion made by another participant, which is not aligned with the agent belief model. In this case, the assertion need not be specifically directed towards the agent.

The additional <text> in each of the template below provides further explanation as to why the agent might disagree with the participant.

Templates to choose from:

1. I disagree <person X>. because <text> 2. I don’t think so, <person X>. <text>
2. not sure, <person X>. <text>, 4. but <person X>, have you considered <text>, 5. but <person X>, think about <text>

There are additional dialogue acts under the Statements and Responses category in our DA hierarchy. These are Signal-Non-Understanding, Offer-Commit and Acknowledge. The agent does not respond to these types of utterances. A Signal-Non-Understanding signifies
cases where a participant indicates they have not understood what has been said. For example, a participant may use “huh?” or “what do you mean?”. However, the cause of non-understanding may not be clear. Since the agent does not generate any text automatically, all the utterances made by the agent are at least syntactically understandable. Any semantic misunderstanding cannot be repaired automatically. We determine experimentally that the probability of Signal-Non-Understanding type of utterances is very low.

While Offer-Commit utterances are common in task-oriented types of dialogues, they appear with very low frequency in open-ended discussions (Stolcke et al., 2000). Accordingly, we did not program an agent response for this type of utterance. An Acknowledge type of utterance signifies that a participant has seen or read what has been presented in chat, and does not typically require a response, according to conversational norms (Webb, Benyon, Hansen and Mival, 2010).

The main reason for not programming behaviors in response to a subset of dialogue act tags is to gain efficiency in rapidly assigning the DA categories for the agent to react appropriately in real-time conversation. In making this decision, we choose speed over meticulousness, responsiveness over thoroughness.

In this manner, we have programmed a set of behaviors in the agent to engage in conversation to a number of different situations. The above does not imply that the agent is simply reactive in nature, only responding to situations initiated by other human participants. We have included in the agent a set of behaviors that are proactive (Section 4.6); these are a sequence of steps that are engaged at different waypoints in the conversation. The Chat Context builder in the Chat Analyzer module sets these waypoints.
Additionally, if the Belief Modeler module sends a signal that all belief models of all participants have sufficiently aligned with the agent belief model on a certain theme, the agent will no longer make arguments in favor of that theme and begin persuasion on a different theme. The utterances in these sequences are of the type Make Assertion Opinion behavior (c.f. Section 4.5.4). We shall describe how these proactive behaviors are programed in the agent in the next section.

4.6 Proactive Behaviors

Before the conversation begins, the algorithm has access to the belief models of the participants. This information is obtained from their pre-discussion survey responses. The agent’s stance on the topic is also known to the algorithm prior to discussion; it is the extreme minority opinion. In order to pursue its persuasive strategies, the system imports the list of arguments on the themes related to the topic (described in Section 3.10). The themes are pre-arranged in order of priority using the following two criteria: 1) the themes that are identified as common targets of persuasion across participants, based on their belief models (these are the themes where the belief strength and belief evaluation will be targeted by the agent); and 2) the themes that have the most arguments associated with them elicited during the Belief Elicitation study. During the conversation, the system accesses those themes from the list of themes related to the topic that are favorable to the position the agent is advocating, and uses the Make Assertion Opinion Statement type of behavior to create utterances to send to conversation when it is acting in proactive mode. The themes that are on the opposing stance of the topic are considered “detrimental” to the agent’s position. We shall expand more on these detrimental themes in the next section, since they play a crucial role in the counter-persuasion strategies implemented in the agent.
As the conversation progresses, the agent uses signals from the Chat Context Builder and the Belief Modeler to determine its behaviors. In the beginning, when the agent receives a signal that the conversation-opening-complete waypoint has passed from the Chat Context Builder, it immediately sends an utterance on the first theme in its list (in the form of Make Assertion Opinion Statement) to introduce the first topic of conversation. Then, the agent switches to the reactive mode and responds to the utterances made by the participants. While the agent reacts and responds, it does not attempt to leverage any participant's response to its own advantage (with the possible exception of the Agree with Reinforcement behavior, where it uses the agreement utterance from a participant as a means to reinforce its own views). The agent might be more effective if it used more strategies that allow it to “play off” the participant’s responses; in the current implementation of the algorithm, we did not program such behaviors. However, we intend to add such capabilities to the agent as part of future work.

The agent switches to proactive mode when it either receives a signal from the Chat Context Builder that all arguments on a theme have been utilized (a waypoint has been passed) OR it receives a signal from the Belief Modeler that the belief model of the participants have changed sufficiently in the intended direction. For example, if it is determined that the belief strength of all participants is on the side the agent is advocating for, then the agent switches back to proactive mode and introduces a new theme. Or, if the belief strength of a participant is detected to move away from the agent in response to a two separate utterances by the agent, the Belief Modeler sends a signal to the Behavior Selector to switch to a different theme. The agent behaves analogously when changing belief evaluation, with respect to signals from the Belief Modeler. In cases when a theme is left unexplored (that there are
arguments still remaining to be made on the theme) based on signals from the Belief Modeler (either indicating success or failure), the theme is flagged for the agent to cycle back in future discussion, but only when other themes have been exhausted. In this manner, the agent works in proactive and reactive modes to achieve its persuasive strategies. In doing so, the agent attempts to exert a high degree of Topic Control, by introducing as many topics as it can. Nonetheless, the participants are not precluded from introducing their own topics. When the agent detects that a participant has introduced a new topic, it will attempt to steer the conversation back to the prior theme that was being discussed.

In the preceding sections 4.5 and 4.6, we described the behaviors that we programmed in the agent. These behaviors enable the agent to deliver persuasive messaging during the conversation and reacting appropriately to the participant’s utterances. In addition to being persuasive, the agent does explicitly need to contend with persuasion from the participants themselves, in the form of counter-persuasive behavior. We describe the counter-persuasion strategies in the next section.

**4.7 Counter-persuasion strategies**

For each topic of discussion, the agent has access to a list of arguments on the themes related to the topic. These arguments are categorized as being either in favor or being unfavorable to the topic. We have previously described the process of creating the list of arguments in Chapter 3 (Section 3.10), and we shall further elaborate upon them here. In Table 4.7.1, we show a fragment of the themes that are imported by the agent at the beginning of conversation. The data presented in Table 4.7.1 are on the topic *GMF: Genetically Modified Foods*. Suppose the agent’s task is to persuade *in favor of* GMF; then all the themes in the Table 4.7.1 that are labeled Pro and their related arguments are considered favorable. All the
themes listed as Anti are negative towards the topic are considered detrimental. If the agent’s task was instead to persuade against GMF, then the arguments labeled Anti would be the favorable arguments to make, while the Pro arguments would be considered detrimental. During the conversation, the agent accesses only the favorable themes and their corresponding arguments to make utterances in the form of the behaviors we have described in previous sections.

<table>
<thead>
<tr>
<th>Position</th>
<th>Theme</th>
<th>Argument Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro</td>
<td>GMFs reduce starvation</td>
<td>People are dying from hunger and (GMFs) could prevent it.</td>
</tr>
<tr>
<td>Pro</td>
<td>GMFs reduce starvation</td>
<td>(GMFs) are able to decrease the costs of foods and reduce the number of those affected by starvation around the world.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(GMFs) are better able to grow in harsher environments and therefore reduce the risk for crops being destroyed in seasons with harsher temperatures or less rain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GMF’s are resistant to conditions that would otherwise impair their ability to grow like weather conditions</td>
</tr>
<tr>
<td>Anti</td>
<td>GMFs have a harmful effect on health</td>
<td>(GMFs) can cause health problems for people later on in their life.</td>
</tr>
<tr>
<td>Anti</td>
<td>GMFs have a harmful effect on health</td>
<td>(GMFs) can cause a number of health problems since these foods aren’t produced by nature, rather by man</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(GMFs) are fake. Food comes from the ground, for as long as humans have been humans we have gotten our food source from the ground, or from other animals, the idea of getting food from a test tube is disheartening.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(GMFs) are not natural. We were not supposed to eat modified food, our bodies are adapted to eat seasonally and fresh food.</td>
</tr>
</tbody>
</table>

Table 4.7.1. Fragment of preprogrammed text for themes/arguments to be used during conversation
As can be seen in Table 4.7.1, there are no arguments included which would allow the agent to counter any beliefs expressed during discussion on the detrimental themes. For example, if the participants discuss the theme *GMFs are fake or unnatural* during a conversation where the agent’s task is to persuade in favor of GMF, there are no arguments that the agent can make to persuade that GMFs are not fake or unnatural.

Detrimental themes are included on the pre- and post-discussion surveys, and therefore, participants are aware of all themes when they make their determinations regarding their opinions on the topic of discussion. Ajzen and Fishbein (1980) recommend that the 10 to 12 most frequently expressed themes elicited from the Belief Elicitation study be included in the discussion surveys. Accordingly, all the themes, which were the 10 to 12 most frequently themes expressed during Belief Elicitation study, were included in the discussion surveys. An equal number of questions on favorable and detrimental themes were presented in the discussion surveys. We wanted to include all the themes related to the topic in the survey instruments in order to obtain a comprehensive representation of the participants’ attitudes.

Now, since the participants are aware of the detrimental themes, they may discuss these themes during conversation. The effect of discussing such themes could be attitude change away from the agent. Moreover, there could be additional themes that are introduced by the participants (not part of the pre-determined list of themes available to the agent), which could also have a detrimental effect if they are discussed during conversation. Accordingly, our counter-persuasion strategies are focused on adding behaviors that enable the agent to contend with the detrimental themes.

The initial counter-persuasion strategies programmed in the agent were quite rudimentary. If the agent detected that a detrimental theme was being discussed, it would continue to
make utterances on the (favorable) theme that was being discussed previously or introduce a new favorable theme. The intention was that by choosing to make any utterances on a different theme, the agent could shift the focus of conversation away from the detrimental topic. Following our initial round of validation experiments detailed in the following chapter, we substantially revised the counter-persuasion strategies described here. We will discuss the revised counter-persuasion strategies in the next chapter.

4.8 Summary of this chapter

In this chapter, we provided the details of the agent algorithm and its implementation. We described the three major modules of that work together for the agent to fulfill its persuasion goals. We explained how the utterances in chat are analyzed on various dimensions automatically. We also represented the belief models and described how they are created and updated. In Sections 4.5-4.7, we presented the behaviors that the agent is equipped with.

With the implementation of the agent thus completed, we then ran experiments to test the efficacy of the agent under various settings. These validation experiments are the subject of the next chapter in this thesis.
Chapter 5: Evaluating the Agent

5.1 Introduction
The aim of the validation experiments was to deploy the agent and validate our persuasion and counter-persuasion strategies in synchronous online conversation environments. Survey instruments administered before and after exposure to the agent in chat conversation provide the core empirical data to evaluate the changes in belief evaluation and strength of the participants, as well as participants’ reactions to other participants, including the agent.

In this chapter, we will describe the design of the experiments in Section 5.2. In Section 5.3, we provide details about the subjects who participated in our experiments. Section 5.4 describes the experimental protocol administered in our experiments. The quantitative and qualitative analysis of data and the findings are presented in Sections 5.5 and 5.6. The summary of this chapter is given in Section 5.7.

5.2 Experiment Design
We designed our experiments in two series. The first of the series was Wizard of Oz (WOZ) experiments that included a human in the loop, acting as the computer agent. The second series of experiments involved the agent acting autonomously.

5.2.1 Wizard of Oz and Autonomous Agent
The Wizard of Oz technique was initially used to develop natural language interfaces (Chapanis, 1982; Good et al., 1984). The basic idea is that a human participant thinks she is interacting with a computer, when in fact she interacting with a human. The primary purpose of the method is to test and refine a computer interface prior to implementation.
One classic use-case is that of understanding user input like speech. Gould et al. (1982), who pioneered the method, simulated an imperfect listening typewriter to find out whether it would satisfy people used to giving dictation. In Gould’s experiment, subjects would speak into a computer microphone. Their task was to dictate letters, such as job applications. A skilled typist, hidden from the subjects, would listen and type the text. An algorithm would intervene before the typed text was displayed to the subject, and make changes to the typed text to simulate different classes of recognizers. For example, the algorithm would filter the typed words against a lexicon of 1000 words, or 5000 words, or unlimited vocabulary. Another class of recognizer simulated “discrete” speech input, where subjects were forced to take unnatural pauses between the words they spoke. From the data collected in the experiment, Gould et al. could determine that user satisfaction increased with increase in recognition vocabulary, and that smaller vocabulary was found to be more frustrating than forced “discrete” speech input.

In this manner, once an application is designed, WOZ studies can help provide opportunities to test and refine the application based on interactions with prototypical users involved in prototypical use-cases (for detailed discussion, c.f. Dahlbäck, Jönsson and Ahrenberg, 1993). In more recent research, WOZ experiments have been used to prototype action games based on computer vision (Höysniemi, Hämäläinen and Turkki, 2004; Henderson et al., 2005). These games involve the use of body movements and gestures to control avatars on screen. Instead of building a functional prototype using computer vision, which might be laborious at the outset, WOZ studies help determine the most intuitive movements for game controls and evaluate the relationship between player actions and avatar actions. Our main motivation to conduct WOZ experiments was to determine the
efficacy of a human agent, the wizard, attempting to persuade individuals in a conversational group setting. If the wizard were successful, it would give a measure of credence to the persuasion algorithm. Additionally, we could analyze the data from these experiments and refine the agent algorithm based on the findings. On the other hand, if the wizard were unsuccessful in persuading participants, we would need to closely scrutinize the underlying causes of such failure, and reengineer the algorithm or revise the experimental protocol before experimenting with an autonomous agent. Thus, our WOZ experiments were designed so that the participants were not informed apriori about the presence of the wizard in the conversation. This makes our WOZ similar to existing WOZ studies (Höysniemi, Hämäläinen and Turkki, 2004; Höysniemi and Read, 2005). In our case, if the participants were made aware of the purpose of study, that a computer agent was one of the participants and that it had as a task to persuade others towards its point of view, the validity of the persuasion manipulation would be compromised.

In the second series of experiments, the persuasion agent acted autonomously throughout the discussion, in that, there was no human wizard controlling the output of the agent. We did not inform the participants of the presence of a persuasion agent in the chat room in these experiments either. We have discussed the rationale behind this ostensible deception in detail in Chapter 1, when we discussed the challenges facing the agent (specifically Challenge 1: The agent must remain undetected as a computer program).

5.2.2 Small Groups and Majority-Minority Influence

We conducted controlled experiments whose design included a small number of human participants and the agent in an online synchronous chat room. We selected the majority-minority influence setting (Gardikiotis, 2011) in our experiment design. In social influence
research, it has been found that “minorities influence people’s thinking, attitudes, and behavior by being consistent in their views and flexible in their negotiation with majority members” (Gardikiotis, 2011). Moscovici, in his empirical work (1969), stated that majorities are often unconcerned about what minorities think about them, minority influence in such cases is rarely based on normative social influence. Normative social influence is defined as the need to conform to the accepted social norms and views of the larger group (Aronson, Wilson and Akert, 2005). Instead, minority influence is usually based on informational social influence - providing the majority with new ideas and new information, which leads them to re-examine their views. Providing new ideas and new information to reexamine views is highly compatible with our chosen persuasion model. Studies involving minority influence conducted by social psychologists (Crano, 2012; Martin and Hewstone, 2008; Nemeth, 2012) also suggest that systematic processing of information is more likely to occur if the minority is consistent in their views. Additionally, Martin, Martin, Smith and Hewstone (2007) found that minority-induced attitude change is more likely to translate into consistent behavioral action by the group than majority-induced attitude change. Specifically, Martin, Martin, Smith and Hewstone (2007) showed that minority influence leads to systematic processing of the persuasive message content, resulting in stronger attitudes.

In terms of group size, Thomas and Fink (1961) and Saegert (1971) who reviewed numerous studies found that as group size increases, the strain on members to organize exchanges also increases. Moreover, as group size increases, so does the diversity in the attitudes held by individuals in the group (Pratkanis, Breckler and Greenwald, 2014). They also found that a larger group size also leads to the possibility of sub-group formation.
Blumberg (2008) and others consider a small group as typically one that can engage in ‘face-to-face’ social interaction or its virtual equivalent and suggest a group size range from 2 to 7 members. Classically, James (1951) established that groups tend to gravitate towards the smallest size, suggesting that larger groups are indeed molecular arrangements of smaller groups. Fairhurst and Chandler (1989) studied interactions between one leader and three members in a group to investigate how they display power and social structure through their use of language. Studies in computer-mediated communication such as those by Bazarova, Walther and McLeod (2012) investigated minority influence in virtual groups of four participants. Accordingly, we chose a group size of 4 in our experiments, which is the minimum size required in order to obtain a majority-minority influence setting. The small group size was pragmatically motivated by the need of assuring reasonably swift response times for the agent and the computational complexities of keeping track of real-time conversation across all participants. Each chat session in our experiments consisted of four participants, two majority opinion holders, one minority opinion holder and the agent (either wizard or computer agent). In order to be consistent across all chat sessions, and also to anonymize the data, we selected pseudonyms for each participant.

In Table 5.2.1, we show the participant identifiers and their role in the conversation. The identifiers were chosen because they are gender-neutral⁴ to ensure that any gender biases would not further confound our experimental findings. The agent was given the identifier zen, and the minority opinion holder was given the identifier memphis. The two majority opinion holders were given the identifiers star and river. The id river was given to the participant who held a more extreme view or was more misaligned with the agent model.

⁴ http://nonbinary.org/wiki/Names
than the participant who was given the identifier *star*. These identifiers were kept consistent across all chat sessions, so that in each chat session the participants *star* and *river* were the majority and *memphis* was the minority. The determination of participant identifiers and their roles was decided from their pre-discussion survey responses on the topic, and consequently, their belief models.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Role</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>zen</td>
<td>Agent</td>
<td>Extreme, in favor</td>
</tr>
<tr>
<td>memphis</td>
<td>Minority opinion holder</td>
<td>In favor, more closely aligned with agent model</td>
</tr>
<tr>
<td>star</td>
<td>Majority opinion holder</td>
<td>Not aligned with agent model</td>
</tr>
<tr>
<td>river</td>
<td>Majority opinion holder</td>
<td>Not aligned with agent model (to a greater extent than <em>star</em>)</td>
</tr>
</tbody>
</table>

Table 5.2.1 Participant identifiers and their role in conversation

Two participants, whose belief models were in opposition to the agent belief model, formed the majority. Only one participant, whose belief model was more or less aligned with the agent belief model formed the “minority”, along with the agent. We explain this using Figures 5.2.1-5.2.3. In Figure 5.2.1, we show the belief model of the agent arguing in favor of topic MLDA (*Lowering the minimum legal drinking age from 21 to 18*). In our experiments, the belief model of the “minority” is not completely aligned with the agent model. As such, the minority (*memphis*), while more closely aligned with the agent than *river* or *star*, also needs to be persuaded on certain facets of the topic. The agent task is to persuade even *memphis*, the so-called “minority” on certain themes related to the topic, while ensuring that those themes on which *memphis* already holds a favorable position are not changed to the opposing view. Only the agent belief model is characterized by the consistent and unchanging stance on the topic as advocated by proponents of the minority influence model (Moscovici, 1980).
Figure 5.2.1. Belief model of the agent persuading in favor of the topic MLDA

Figure 5.2.2. Belief model of a participant holding majority opinion on topic MLDA, not aligned with agent belief model shown in Figure 5.2.1
In Figure 5.2.2, we show the belief model for a participant who holds the majority opinion and in Figure 5.2.3, we show the belief model of a “minority” opinion holder. As we can see, there can be differences as well as consistencies between the participants and the agent, illustrated both in Figure 5.2.2 and 5.2.3. Crucially, the attitude of the participant holding the majority opinion is not aligned with the agent model, while the minority opinion holder is more closely, albeit only partially, aligned with the agent model.

We provide information related to the participants and experimental schedule in the next section.

5.3 Participants and Schedule

Subjects participating in both series of experiments were recruited via in-class recruitment. Subjects were compensated through extra credit of up to 3 credits at the discretion of faculty who identified willing students for this study. All necessary institutional review board approvals were obtained prior to subject recruitment.
In Table 5.3.1, we provide demographic details of participants who participated in the WOZ and autonomous agent (AA) experiments. A total of 22 participants were recruited for the WOZ experiments, of which 18 (82%) participated to completion. The sample was comprised of 13 female and 4 male participants, and 1 participant who did not wish to disclose a gender. The WOZ experiments were conducted over a period of three weeks between November 23rd and December 11th, 2015. Six 60-minute chat sessions were conducted, each involving 4 participants, of whom 3 were the recruited subjects and 1 was the human wizard. Four participants who did not participate to completion included 2 subjects who expressed interest in participating but did not complete the pre-discussion survey when we emailed them and 2 subjects who completed the pre-discussion survey but later emailed that they would not be able to come to the laboratory for the discussion session.

After a phase of data analysis and algorithm refinement (described in Section 5.6) following the WOZ study, we began our next series of experiments. These were undertaken between March 21st and April 7th, 2016. 19 participants were recruited for the autonomous agent experiments. 12 participants (63%) completed the study in its entirety. 7 male and 5

<table>
<thead>
<tr>
<th>Demographic</th>
<th>WOZ experiments (n=18)</th>
<th>Autonomous experiments (n=12)</th>
<th>Across both experiments (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>21.17</td>
<td>22.2</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.82</td>
<td>2.36</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female (%)</td>
<td>13 (72%)</td>
<td>5 (42%)</td>
</tr>
<tr>
<td></td>
<td>Male (%)</td>
<td>4 (17%)</td>
<td>7 (58%)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Caucasian (%)</td>
<td>11 (61%)</td>
<td>6 (50%)</td>
</tr>
<tr>
<td></td>
<td>Non-Caucasian (%)</td>
<td>7 (39%)</td>
<td>6 (50%)</td>
</tr>
</tbody>
</table>

Table 5.3.1. Demographic details of participants who participated in our validation experiments
female participants took part in this study. As with the WOZ chat sessions, each 60-minute chat session involved 3 recruited participants and 1 computer agent conversing in an online chat room. The participants from WOZ experiments were not eligible to participate in the AA experiments.

The average age of participants across both studies was between 21-23 years. The ethnicity of participants in the WOZ experiments was 61% Caucasian, and 50% in the autonomous agent experiments. Out of the 19 subjects who expressed interest in participating in the second phase of the study, 4 did not complete the pre-discussion survey, and 3 participants completed the survey but did not come to the laboratory for the discussion session. In total, 30 participants were part of the validation experiments across both the WOZ and autonomous settings.

After describing our recruitment procedures and experiment schedule, we now describe the experimental procedure, including participant laboratory experience and debriefing.

5.4 Procedures

Procedures for both the WOZ and AA experiments were nearly identical. The only difference was the presence of the wizard as one of the participants during the WOZ experiments, whereas, in the AA experiments, the algorithm acted autonomously. As the first step, participants were asked to complete a pre-discussion survey. This survey was administered online. Participants were sent a link to survey website in their email. Participants were also sent two reminders evenly spaced over a week, if they did not complete the survey upon receipt of email. On the first screen of the survey, participants were presented the consent form, which they had to read and sign before moving to answering the survey questions. The next screen presented them with a brief set of
instructions regarding how to answer the survey questions (Appendix B: copy of pre-
discussion survey). Once the participants read the instructions, they proceeded to
answering belief evaluation and belief strength questions regarding each of the six topics
we have chosen for our research purposes (c.f. Chapter 3, Section 3.4). The questions
regarding belief strength and evaluation were set up according to the procedure we
described in Chapter 3. Questions for each topic were blocked together, that is, all belief
strength and evaluation questions regarding the topic Genetically Modified Foods (GMF)
were presented on a single screen. For example, one set of questions on the GMF topic is
presented in Figure 5.4.1. For each topic, there are such sets of questions, related to the
themes we selected to include in the surveys (Section 3.9) presented on a single screen,
followed by the question shown in Figure 5.4.2.

How likely do you feel the following statement to be true?

Genetically modified foods have a harmful effect on health:* 
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided  
( ) Somewhat likely ( ) Likely ( ) Very likely

Having a harmful effect on health is:* 
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad  
( ) Fair ( ) Good ( ) Very Good

Figure 5.4.1. Questions assessing belief strength and belief evaluation on our survey

The order of topics was randomized for each participant, to overcome the effect of
response fatigue across participants. At the end of the belief strength and evaluation
questions for each topic, participants were asked to answer a question assessing their
overall attitude towards the topic (shown in Figure 5.4.2 for topic GMF).
In addition, there were two questions that assessed the participants’ level of knowledge and their willingness to participate in a discussion for each topic. These questions are presented in Figure 5.4.3 and 5.4.4. The last question on the survey asked participants to indicate their availability for any of the scheduled chat sessions.

**How informed are you about the following topics? Please rank each topic in order (1=most informed, 6= least informed). You can only choose one response per topic:**

*Figure 5.4.3. Question assessing how informed a participant feels about the discussion topics*

**How interested are you in discussing the following topics? Please rank each topic in order of your level of interest (1=most interested, 6= least interested). You can only choose one response per topic:**

*Figure 5.4.4. Question assessing participant interest in the discussion topics*

### 5.4.1 Participant laboratory experience

We scheduled chat sessions with the participants, by composing the groups based on the following criteria, not necessarily in order of importance:

1. Availability
2. Level of interest in discussing topic
3. Level of knowledge on topic
4. Ability to fill in role of majority or minority position on the topic based on their survey responses on belief strength and evaluation questions.
As can be noted, many different factors are at play when deciding which participant should be scheduled for which session.

Participants were sent an email informing them to come to the laboratory at the appointed date and time after we organized the sessions based on the above criteria. Each participant was also sent an email reminder the day before the chat session. The experimenter greeted each participant upon arrival. They were given their signed consent form to review and informed of their rights afresh. They were asked to take note of their assigned chat identifier (river, star or memphis) and were logged into a secure online chat room. Participants were only made aware of their own identifiers. They did not know which identifiers were assigned to which other participants.

Figure 5.4.1.1. Screenshot of the chat interface
In Figure 5.4.1.1, we show a screenshot of the chat interface, with some sample messages. Participants type their utterances in the bottom-left window (labeled Message) and press the Enter key to send their utterances to the chat room. All messages sent by participants in the chat room are displayed in the top-left window.

Once all 3 participants scheduled for chat session arrived in the laboratory (as well as the wizard in case of WOZ experiments), we commenced the chat session. Participants were collocated in the same room, at different computer stations. The computer stations were arranged so that participants were facing away from each other. Additionally, there were cubicle partitions arranged so that participants could not see the others during the discussion. As explained above, in neither experiment did we inform participants that one of the discussants would have a hidden purpose during conversation. In the autonomous agent experiments, we informed subjects that one of the participants would be participating from a remote location, although we did not reveal the identifier of that participant (zen).

Participants were told which topic they would be discussing, for a period of one hour. No additional instructions were given. They were free to converse in a spontaneous fashion without any limitations or restrictions placed on the content of their discourse. The experimenter tracked the discussion on her own screen and also observed the participants unobtrusively to make notes regarding any significant events during the session.

At the end of one hour, the experimenter sent a message from her screen to ask the participants to end the discussion. She also sent a link to the post-discussion survey and asked participants to complete the survey in the laboratory.
The post-discussion survey (Appendix C) contained belief strength and evaluation questions for a single topic, the topic that the participants had presently discussed. The post-discussion survey also contained the same question assessing the overall attitude towards topic, as shown in Figure 5.4.1.

During the discussion, some people are more persuasive than others. Below is a list of participants, including yourself. Please rate each participant in terms of how persuasive they seemed to you:*

![Figure 5.4.1.2. Question assessing peer ratings of influence](image)

Additionally, there were questions to ascertain the impact of the agent during conversation. We measure impact through the questions 5.4.1.2 and 5.4.13. We show these in Figures 5.4.1.2 and 5.4.1.3. In Figure 5.4.1.2, we present the question assessing peer ratings of persuasion. Each participant (not including the agent) would rank themselves and their fellow participants on their degree of persuasion during the conversation. In addition, we added a question specifically concerning the persuasiveness of the agent, as shown in Figure 5.4.1.3.

One participant is selected at random from this conversation. Please indicate your opinion for the following question.*

Participant zen was persuasive?

( ) Strongly Agree ( ) Agree
( ) Neither Agree nor Disagree ( ) Disagree ( ) Strongly Disagree

![Figure 5.4.1.3. Question assessing persuasiveness of the agent](image)

After the participants completed the post-discussion survey, we presented them with a debriefing form (Appendix D: Copy of debriefing form) and informed them of the real
purpose of the study. We requested participants to not divulge the real purpose of the experiment to anyone, so that future participants would not know about the ostensible deception. Participants were free to ask questions and were also free to ask that their data be removed from analysis. None of the participants who completed the study asked for their data to be removed. We also asked participants if they would be willing to discuss their impressions in interviews with the experimenter. One participant from each session was selected randomly for interviews. Seven qualitative interviews were conducted across both experiments, which represent 23% of the sample.

The completion of post-discussion survey and following interview, if any, ended the subject participation in the experiment.

For the WOZ experiments, we asked a human wizard to play the role of the persuasion agent. In order to ensure the validity of the experiment and utility of the data resulting from chat sessions involving the wizard, we extensively trained the wizard to perform their role appropriately. In the following section, we explain the wizard training procedure in detail.

5.4.2 Wizard training

It was imperative that the wizard be suitably trained in order to perform their assigned role in conversation. A graduate student from the Communication Department of our university was selected to be the wizard. The student was quite familiar with the Summative Model of attitude change, having studied social influence as part of their coursework. The student also had a Bachelors degree in Computer Science, which gave him an additional appreciation of the intricacies of interacting with a computer program built to persuade automatically.
The wizard underwent a series of training exercises prior to participating in actual discussions with recruited subjects. Initially, the experimenter and the wizard met for two 1-hour discussions regarding the specifics of experiment design and the persuasion algorithm. During the training discussions, the wizard was made familiar with the chat interface and the support interface for the wizard. In Figure 5.4.2.1, we show the support interface that was designed for the wizard.

As shown in Figure 5.4.2.1, the algorithm presents up to three options to the wizard. In the example shown above, the topic is MAP: Easy access to morning after pill to people under 17. The utterances presented by the algorithm are Make Assertion Opinion Statement type of behavior. The agent’s goal in this example is to persuade in favor of topic MAP. The wizard
may select one or none of the options presented by the algorithm. In this case, the algorithm presented the agent with two options. The wizard may modify the text in any of available options before selecting it to be sent to the chat discussion. The wizard was instructed to change the text slightly, only if there were any spelling or grammar issues found. The extent of flexibility given to the wizard with respect to changing the messages produced by the algorithm was deliberately limited, because we wanted to ascertain the effectiveness of the algorithm as accurately as possible under controlled conditions. If the wizard were allowed more freedom, it would become nearly impossible to isolate the effects of persuasion under such conditions.

Clicking on one of the buttons labeled ‘Use this!’ would send the corresponding text to the ongoing chat discussion. The other participants would view the text as appearing from participant zen.

During the training sessions, the wizard and the experimenter engaged in mock discussions while using the chat and support interfaces, so that the wizard would become comfortable with using the interfaces and responding to various types of situations. In addition, there were two training sessions, where the wizard engaged in mock chat sessions with three other participants to get acquainted with reacting to the dynamics of conversation with the same number of participants as would be present during actual chat sessions. The three other participants in these mock sessions were research personnel affiliated to our laboratory.

The wizard was compensated financially for 20 hours of his time. In all, the wizard underwent 10 hours training and spent 10 hours participating in 6 chat discussions during the WOZ experiments, including preparation and post-discussion debriefing. The wizard
debriefing was conducted after all the participants had left the laboratory. Brief qualitative analysis of wizard debriefing is presented in Section 5.5.4.

In the next section, we describe the measures and metrics used in our study.

### 5.4.3 Measures and Metrics

Our main research hypothesis is that the interventions made by the agent would result in attitude change in the participants. Pre- and post-discussion surveys allow us to measure changes in participant’s belief models, and thus, the shifts in their overall attitude towards the topic of discussion. Pre- and post-discussion surveys have been used in a variety of contexts to measure intervention outcomes (Price and Capella, 2002; Goncalo and Staw, 2006; Price, 2009; inter alia). Andrews, Mandahar and DeBoni (2008), who used belief change as a measure of persuasion in their work on a persuasive computer dialogue system, based their persuasiveness measure on the “Kendall τ permutation metric” (Kendall & Gibbons, 1990). The Kendall τ metric is used to compute the pairwise disagreement between two rankings, in that, it measures the number of swaps needed between adjacent items in one ranking to obtain the target ranking. Andrews et al. applied this metric in their experiments where the system goal was to persuade participants to change their ranking on a list of items, for example, in the Desert Survival Scenario (Moon, 1999). In the experiments conducted by Andrews et al., the system’s task was to persuade humans to change the ranking of items they would salvage from a plane if they had crashed and were stranded in a desert. The persuasiveness measure using the Kendall τ metric then measures the number of swaps needed from the subject’s pre-discussion ranking to obtain their post-discussion ranking, after the subject has undergone a discussion session with the persuasive system. The system goal in our experiments is more complex, in that, our
persuasion algorithm is multifaceted, attempting to change the belief strength and belief evaluation, and furthermore, the overall attitude of the receivers of persuasive messaging. Accordingly, we use the Wilcoxon signed-rank test (Wilcoxon, 1945) to measure persuasion effectiveness of the algorithm, which is a test to compare two matched samples, as an alternative to the paired t-test. In particular, it is suitable for evaluating data from a repeated-measures design, where the prerequisites for a dependent samples t-test, such as normally distributed samples, are not met. For example, the Wilcoxon signed-rank test has been used to evaluate the reading ability of children before and after they undergo a period of intensive training (Sim, MacFarlane and Read, 2006). In our case, the Wilcoxon signed-rank test is used to evaluate changes in overall attitudes of participants before and after (these form the matched samples) they are subjected to interventions during the chat discussion.

In addition, we use peer-ratings of participants to indicate the ranking of persuasiveness. Self-assessment and peer-assessments of behaviors have been used in a variety of studies (Turentine, 2001; Charbonneau, 2002; Strzalkowski et al., 2010, inter alia). We specifically asked participants to rank themselves and their peers on their degree of persuasiveness. In order to further substantiate persuasion rankings, we obtained automated rankings of influence using the DSARMD system (Strzalkowski et al., 2010; 2012). The DSARMD system provides a ranking of all participants in conversation on complex sociolinguistic phenomena such as leadership, influence and power. The system has been validated against human assessments of such phenomena and has been shown to have an accuracy of >90% across the various sociolinguistic phenomena and discussion types (synchronous chat conversation, asynchronous discussion forums, etc.) (for an extended discussion, c.f.
Strzalkowski et al., 2010; 2012). We use the DSARMD system to provide convergent evidence of our agent’s rank on influence, and other sociolinguistic behaviors such as involvement. We use the automated DSARMD system ranks to augment the peer and self-assessments made by the participants of conversation.

The rating measures above constitute our metrics for validating the agent. In the next few sections, we present the analysis of data and findings from our experiments.

5.5 Data Analysis and Results – WOZ experiments

We begin with a descriptive analysis of chat sessions, followed by the results on perception ratings, belief change and wizard debriefing interviews. In each section, we will highlight the key findings that we observed in the analyses, which led to the overall conclusions and lessons learnt from these first series of experiments.

5.5.1 Descriptive Statistics

<table>
<thead>
<tr>
<th>Discussion Topic</th>
<th>woz_session_1</th>
<th>woz_session_2</th>
<th>woz_session_3</th>
<th>woz_session_4</th>
<th>woz_session_5</th>
<th>woz_session_6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Being able to easily identify people online</td>
<td>Lowering minimum legal drinking age to 18 from 21</td>
<td>Easy access to morning after pill to people under 17</td>
<td>Lowering minimum legal drinking age to 18 from 21</td>
<td>Being able to easily identify people online</td>
<td>Easy access to morning after pill to people under 17</td>
<td></td>
</tr>
<tr>
<td>Agent position</td>
<td>Pro-topic</td>
<td>Pro-topic</td>
<td>Pro-topic</td>
<td>Pro-topic</td>
<td>Pro-topic</td>
<td>Pro-topic</td>
</tr>
<tr>
<td>Total utterances across all participants</td>
<td>145</td>
<td>201</td>
<td>155</td>
<td>184</td>
<td>178</td>
<td>204</td>
</tr>
<tr>
<td>No. of utterances made by wizard (%)</td>
<td>23 (16%)</td>
<td>39 (19%)</td>
<td>24 (15%)</td>
<td>30 (16%)</td>
<td>29 (16%)</td>
<td>32 (16%)</td>
</tr>
</tbody>
</table>

Table 5.5.1.1 Overview of the six chat sessions conducted during WOZ experiments
Six chat sessions were conducted during the WOZ experiments. In Table 5.5.1.1, we give details regarding these sessions. We name them woz_session_1 through woz_session_6. The topic of discussion and the position held by the agent was predetermined by the composition of the group and their belief models. The average number of utterances across all participants in these chat sessions was 177.83 (stdev=23.9). The wizard made an average of 29 utterances across the sessions, which represents approximately 16% of the discussion content.

In Figure 5.5.1.1, we show the distribution of utterances for all participants in the six WOZ chat sessions. We found significant differences in the number of utterances made by the wizard and each of the participants (river, star and memphis) at p <= 0.05.

We ranked all the participants on their degree of involvement using the automated DSARMD system. The DSARMD system computes multiple indices from textual input to determine the level of involvement of each participant, and number of turns made by the participant is one of the indices of involvement. The other indices of involvement include 1) the proportion of nouns and noun phrases used by the participant; and 2) being part of the discussion on the most persistent topics of conversation. Using the DSARMD system rankings, we ascertained that zen ranked either second or third out of the four participants across all WOZ chat sessions (Table 5.5.1.2). This indicates that although the system made significantly fewer turns than the three other participants, there were other sociolinguistic behaviors such as contributing content words (in the form of nouns and noun phrases) and being part of the discussion on persistent topics that were indicative of involvement of the system.
Figure 5.5.1.1 Distribution of number of utterances for all participants across all WOZ chat sessions

Table 5.5.1.2 Ranks and scores of participants on Involvement behavior computed using the automated DSARMD system

Figure 5.5.2 Distribution of types of behaviors produced by the wizard
In Figure 5.5.1.2, we show the types of behaviors exhibited by the agent across all six WOZ chat sessions. A majority of statements made by the agent were of the *Make Assertion* *Opinion Statement* type. *Disagree with Reasons* and *Agree with Reinforcement* also make a substantive proportion of the remainder of the data. This observation serves as a validation of the algorithm design. In each of these top three behaviors, the agent has the opportunity to assert its own beliefs using natural language text, in addition to the pre-formatted templates. These behaviors were described in Chapter 4, Sections 4.5 and 4.6.

**Finding 1:** The substantial proportion of the top three proactive behaviors indicates that the proactive algorithm strategies to persuade others occur more often than the reactive type of behaviors. The wizard made significantly fewer utterances during conversation when compared to the other participants, however, the degree of involvement of the system was not the lowest in any WOZ chat session and thus would not be considered uninvolved.

**5.5.2 Perception Ratings**

As described in Section 5.4.1, two questions in the post-discussion survey were designed to reveal the participants’ perceptions of the agent’s persuasiveness. We analyzed the responses to the perception question, which directly asked whether they thought *zen* was persuasive. We found that a majority of the participants rated *zen* to be persuasive. These data are presented in Figure 5.5.2.1. Thirteen (73%) participants either agreed or strongly agreed with the statement ‘Participant *zen* was persuasive’. Four participants either disagreed or strongly disagreed with the statement ‘Participant *zen* was persuasive’, which represents 17% of the population.
The ranking question shown in Figure 5.4.1.2 was used to obtain peer ratings of persuasiveness. We found that the average rank given to the wizard on persuasiveness was 2.33, across all sessions (1 being the most persuasive, 4 being the least). We also obtained the automated system rankings from the DSARMD system (Strzalkowski et al., 2010; 2012). The average rank for participant *zen* in these automated rankings was 2. The automated rankings and those given by participants to the agent show a weak positive correlation (Pearson’s *R* = 0.45). The ranking data also revealed that participant *river* was ranked higher than *zen*, both in the peer-ratings by fellow participants (average rank = 1.67) and by the automated DSARMD system (average rank = 1.33). This was not unanticipated given that the stronger majority participant was consistently assigned to be *river*.

**Finding 2:** Although a substantial proportion of participants perceived *zen* to be persuasive, they ranked participant *river* to be more persuasive on average than *zen*. 

![Distribution of responses on question: Participant zen was persuasive](image)
5.5.3 Persuasion Effectiveness

We used the overall attitude score (calculated as the sum of the products of belief strength and belief evaluation across all themes related to the topic $\sum b_i e_i$, please see Chapter 2, Section 2.2.3 and Chapter 4, Section 4.4) from pre- and post-discussion surveys to determine changes in participant attitudes. Results from the Wilcoxon signed rank test for persuasion effectiveness using change in overall attitude from pre- and post-discussion surveys were not statistically significant in the WOZ experiments ($Wilcoxon N=18, p=0.76418$). In Figure 5.5.3.1, we nonetheless plot the change in attitude for all eighteen WOZ experiment participants, and also the mean attitude change (-1.44) and stdev (21.07). We described in Table 5.5.1.1 that the position of the agent was pro-topic in all WOZ chat sessions. Consequently, a positive change in the below graph may be construed as successful persuasion, while a negative change can be attributed to unsuccessful persuasion. On the x-axis, we show all eighteen participants (numbered 1 through 18 for ease of presentation), and on the y-axis we show the change in their attitude from pre- to post-discussion. The results presented in Figure 5.5.3.1 indicate that the system was successful in achieving the intended attitude change in 8 out of 18 participants (44%), where some participants showed a greater attitude change than others, aligning more closely with the agent model than others. The rest of participants were not successfully persuaded (56%). If anything, they showed variability in their degree of attitude change and movement away from the agent model.
We show the changes in attitude of each participant for all six sessions in Figures 5.5.3.2-5.5.3.7 below. Each chat session is shown as a separate chart. The topic of discussion is displayed as the chart title. The participants (*river*, *star* and *memphis*) are named on the x-axes and their attitude totals prior to and after the discussion are shown on the y-axes in these charts. In each chat session, the participant *river* holds the strongest “majority” opinion, followed by *star* and then by *memphis*.

For woz_session_1 (shown in Figure 5.5.3.2), we see that the persuasion is not successful for any participant. As noted before, the position of the agent was pro-topic in all WOZ chat sessions. Consequently, a positive change in the graphs may be construed as successful persuasion, while a negative change is an instance of unsuccessful persuasion. All participants, *river*, *star* and *memphis* in woz_session_1 move away from the position advocated by the agent. A similar observation is made for woz_session_2, shown in Figure 5.5.3.3. We see that although *memphis* still remains on the positive side of the topic, the
change in attitude is away from the agent. In Figure 5.5.3.4 (woz_session_3), we see that the persuasion has been successful for river and star. The change in attitude for both these participants is towards the system’s belief model. In this session, memphis does move slightly away from the agent (total of $+12 \sum b_i e_i$ prior to the discussion, $\sum b_i e_i +10$ after the discussion). Figure 5.5.3.5 (woz_session_4) shows that the persuasion is successful across all three participants, with star showing the most notable movement towards the system (+20 towards the system). However, in this session, the participant star was somewhat positive towards the topic prior to the discussion ($+10 \sum b_i e_i$).
Figure 5.5.3.6 and Figure 5.5.3.7 Change in attitudes for participants in WOZ session 5 and session 6

Figure 5.5.3.6 (woz_session_5) shows unsuccessful persuasion for all three participants. We also noticed that the topic in WOZ session 1 and 5 is the same, viz. *BIO: Being able to identify people online*. We will expand upon this discovery at the end of this section. In Figure 5.5.3.7 (woz_session_6), we see that persuasion is successful for all three participants, with *river* showing the maximum change in attitude (-36 $\sum b_i e_i$ prior to discussion, $+5 \sum b_i e_i$ after discussion, total change of $+41$ towards the system belief model).

From the Figures 5.5.3.2-5.5.3.7, we can consider sessions 3, 4 and 6 to be successful sessions, while the other sessions are unsuccessful. Also, the magnitude of positive change in overall attitude is highest in woz_session_6, followed by woz_session_4 and then by woz_session_3.

We conducted a more detailed analysis of the data to identify the specific causes of success and failure. The persuasion strategies and the themes on which the system might attempt to persuade were pre-programmed in the agent as previously discussed in Chapter 4. We used the responses on individual belief strength and belief evaluation questions related to the overall topic of discussion, to see which strategies achieved the intended results and
which strategies were unsuccessful. Figure 5.5.3.8 shows the distribution of persuasive strategies and their effect over all participants across all sessions. We categorized the changes in belief strength and belief evaluation from pre- to post-discussion surveys in three major categories: 1) successful persuasion in changing the belief strength by at least one point towards the agent belief model when pre-discussion belief evaluation on a theme is aligned with the agent model; 2) successful persuasion in changing belief evaluation at least one point towards the agent belief model when pre-discussion belief strength on the theme is not strong viz. belief strength [-2, 2] values; and 3) unsuccessful persuasion, change in belief strength or belief evaluation which is away from the agent belief model.

![Diagram](image)

**Figure 5.5.3.8** Proportion of instances of attitude change from pre- to post-surveys on all themes across all WOZ chat sessions

Figure 5.5.3.8 shows that for the responses collected during WOZ chat sessions, the proportion of cases where the belief strength or belief evaluation of the participant did not change towards but instead changed away from the agent belief model is 55%; these are the instances of unsuccessful persuasion. There were a small proportion of cases (1%)
where no change in belief strength and belief evaluation on a theme was observed from pre- to post-discussion survey response. We classify these cases in a separate category, since they can be considered neither successful nor unsuccessful persuasion instances. A small proportion of cases (2%) were identified where successful persuasion could not be attributed to weak belief strength or belief evaluation aligned with agent model. We shall expand upon the categories below.

1. The agent was successful in changing the belief strength on themes where the belief evaluation aligned with the agent model in 19% of cases. For example, consider the following responses (shown in Table 5.5.3.1) by participant *river* on one theme related to the topic of *MAP: Easy access to morning after pill to people under 17* taken from *woz_session_3*.

<table>
<thead>
<tr>
<th>MAP leads to fewer teenage pregnancies</th>
<th>bi</th>
<th>ei</th>
<th>bi*ei</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>river</em> pre-discussion response</td>
<td>-3</td>
<td>3</td>
<td>-9</td>
</tr>
<tr>
<td><em>river</em> post-discussion response</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Change</td>
<td>+5</td>
<td>0</td>
<td>+15</td>
</tr>
</tbody>
</table>

Table 5.5.3.1 Example of successful persuasion on theme when belief evaluation aligns with agent model

The data presented in Table 5.5.3.1 indicate an instance of persuasion success on a theme when the evaluation aligns with the agent model. Participant *river* (participant with strong majority attitude) belief strength for the theme *MAP leads to fewer teenage pregnancies* is -3 (very unlikely) prior to discussion, and it changes to +2 (likely) after the discussion resulting in overall +15 change on that particular theme. We see that a relatively small change in belief strength (+5) can lead to a comparatively large difference in overall attitude (+15 in favor of the agent)
It is not necessary that the belief evaluation be exactly the same as that of the agent. In Table 5.5.3.2, we show another example on the same theme from the same chat session for participant *memphis* (participant with weak minority attitude) where the initial belief evaluation (+2, good) is slightly lower than that of the agent (+3, very good).

<table>
<thead>
<tr>
<th>MAP leads to fewer teenage pregnancies</th>
<th>bi</th>
<th>ei</th>
<th>bi*ei</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>memphis</em> pre-discussion response</td>
<td>-2</td>
<td>2</td>
<td>-4</td>
</tr>
<tr>
<td><em>memphis</em> post-discussion response</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Change</td>
<td>+4</td>
<td>+1</td>
<td>+10</td>
</tr>
</tbody>
</table>

Table 5.5.3.2 Another example of successful persuasion on theme when belief evaluation aligns with agent model

2. In 23% of cases, the agent was successful in changing the belief evaluation when the belief strength was rated weak or undecided. We illustrate such instances with a specific example in Table 5.5.3.3. The example is also taken from the same discussion (woz_session_3) on the MAP topic. In this case, the theme is *MAP allows teenagers more control over their sex lives.*

<table>
<thead>
<tr>
<th>MAP allows teenagers more control over their sex lives</th>
<th>bi</th>
<th>ei</th>
<th>bi*ei</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>star</em> pre-discussion response</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>star</em> post-discussion response</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Change</td>
<td>0</td>
<td>+3</td>
<td>+3</td>
</tr>
</tbody>
</table>

Table 5.5.3.3 Example of successful persuasion on theme when belief strength is weak

Participant *star* (participant with weak majority opinion) does not strongly believe in this assertion and has belief strength of +1 (somewhat likely). Additionally, her belief evaluation is 0 (neither good nor bad) prior to discussion. The agent is able to successfully change her belief evaluation to +3 (very good) during the conversation by
presenting beliefs in favor of allowing teenagers more control over their sex lives, resulting in an overall +3 attitude change on the topic.

3. We observed in 2% of cases there was successful persuasion, however these are cases that cannot be classified in the previous two categories we have listed above. We illustrate via an example in Table 5.5.3.4. The example is taken from woz_session_4 where the topic of discussion was *MLDA: Lowering minimum legal drinking age from 21 to 18*. We present the responses for participant *river* (participant with strong majority attitude) on the theme *MLDA leads to acknowledgment of 18 as age of adulthood*. The belief evaluation of *river* on this theme prior to discussion is neither aligned with the agent model nor misaligned, her belief evaluation is 0 (neither good nor bad). After the discussion, we see a change in both belief strength and belief evaluation, which is in alignment with the agent model.

It can be inferred that in such cases, the successful persuasion occurs because the belief evaluation prior to discussion is neutral, and thus can be changed to strongly positive by the agent, along with corresponding desirable changes in belief strength.

<table>
<thead>
<tr>
<th>MLDA leads to acknowledgment of 18 as age of adulthood</th>
<th>bi</th>
<th>ei</th>
<th>bi*ei</th>
</tr>
</thead>
<tbody>
<tr>
<td>river</td>
<td>-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pre-discussion response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>river</td>
<td>3</td>
<td>3</td>
<td>+9</td>
</tr>
<tr>
<td>post-discussion response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change</td>
<td>+5</td>
<td>+3</td>
<td>+9</td>
</tr>
</tbody>
</table>

Table 5.5.3.4 Example of successful persuasion on theme when belief strength is weak

While the three trends we discuss above indicate instances of successful persuasion, the next trend we discovered reveals the primary cause of failure for the agent.
4. In 55% of the cases, persuasion was unsuccessful. The main reason for attitude change away from the intended belief model were detrimental themes. For detrimental themes, the agent could neither control the topic nor offer any counter-arguments. We discussed detrimental themes in Chapter 4, Section 4.7. We illustrate this trend by two specific examples. In Table 5.5.3.5, we show the change in bi*ei for participant star on the theme *Teenagers may abuse MAP* from the same discussion as presented in examples above (woz_session_3). Prior to the discussion, star does not believe that teenagers may abuse MAP (belief strength -1, somewhat unlikely), but after the discussion the belief strength changed to +1 (somewhat likely). This small shift in belief strength, coupled with a negative belief evaluation leads to an overall change of -5 (away from the intended agent model). The effect of discussing detrimental topics is change in attitudes away from the intended agent model. During the woz_session_3 discussion, it was participant star herself who introduced the theme of *Teenagers may abuse MAP*, which led to the participants (river, star and memphis) discussing it for a sequence of 5 utterances.

<table>
<thead>
<tr>
<th>Teenagers may abuse MAP</th>
<th>bi</th>
<th>ei</th>
<th>bi*ei</th>
</tr>
</thead>
<tbody>
<tr>
<td>star pre-discussion response</td>
<td>-1</td>
<td>-3</td>
<td>3</td>
</tr>
<tr>
<td>star post-discussion response</td>
<td>1</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>Change</td>
<td>2</td>
<td>+1</td>
<td>-5</td>
</tr>
</tbody>
</table>

Table 5.5.3.5 Example of unsuccessful persuasion on theme when belief evaluation is negative and agent offers no counter-arguments

As another example, consider the data presented in Table 5.5.3.6 on the topic *MLDA: Lowering the minimum legal drinking age from 21 to 18*. The theme is *MLDA has adverse effects on developing brains*. For this detrimental theme on the MLDA topic, that agent did
not possess any arguments that would persuade against MLDA leading to adverse effects on developing brains. As a result, when the discussion in woz_session_4 shifted to this particular theme, the agent could not offer any persuasive arguments. Table 5.5.3.6 shows the change in attitude on this theme for all three discussion participants. Participant river changed their belief strength from +1 (somewhat likely) to +2 (likely), resulting in an overall -3 shift away from the agent. Participant star did not change either their belief evaluation or strength on this theme. More noteworthy is the belief strength change in participant memphis. Prior to discussion, she was undecided (belief strength=0) on this theme, and after the discussion she strongly believed this theme (belief strength=3, very likely). This causes an overall -6 shift away from agent belief model. The topic of MLDA has adverse effects on developing brains was introduced by river during the conversation.

<table>
<thead>
<tr>
<th>MLDA has adverse effects on developing brains</th>
<th>bi</th>
<th>ei</th>
<th>bi*ei</th>
</tr>
</thead>
<tbody>
<tr>
<td>river pre-discussion response</td>
<td>1</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>river post-discussion response</td>
<td>2</td>
<td>-3</td>
<td>-6</td>
</tr>
<tr>
<td>star pre-discussion response</td>
<td>1</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>star post-discussion response</td>
<td>1</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>memphis pre-discussion response</td>
<td>0</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>memphis post-discussion response</td>
<td>3</td>
<td>-2</td>
<td>-6</td>
</tr>
</tbody>
</table>

Table 5.5.3.5 Result of unsuccessful persuasion on theme when belief evaluation is negative and agent offers no counter-arguments

We found that when the discussion ventured into themes and topics for which we had not anticipated programming arguments in the agent, the net effect was unsuccessful persuasion on the overall topic. It was not the case that the agent algorithm was ‘silent’
during the periods when detrimental topics are discussed. As described in Section 4.7 in Chapter 4, our rudimentary counter-persuasion strategies focused on attempting to continue the conversation on topics that were beneficial to the system. The algorithm would continue to offer arguments and continue with its own persuasion strategies. However, the utterances made by the wizard/agent during those periods would either be ignored or would not have the desired effect, since the participants were ostensibly interested in discussing other topics. The trends discussed above lead us to our third finding.

Finding 3: Although a sizable proportion of cases were found where persuasion was successful across all six WOZ sessions (45%, c.f. Figure 5.5.3.8), the magnitude of the changes in attitudes of participants was not statistically significant. When the discussion centered on detrimental themes, on which there were no robust counter-persuasive strategies programmed in the system, the effect was belief change in the opposite direction than that desired by the agent. While this finding may not indicate an inherent flaw in the underlying algorithm, we presume that the agent would be deemed even more successful if there were specific mechanisms to cope with the appearance of detrimental topics.

5.5.4 Wizard debriefing results

The findings presented in the sections above were echoed by the wizard during post-discussion debriefing sessions. The wizard noted that he ran out of arguments or was not offered any appropriate choices by the algorithm when the discussion centered on topics detrimental to the agent’s belief model. The wizard also noted that such topics were typically introduced by the majority opinion holders, river and star.
We previously stated that the wizard had an option to remain silent when he deemed that none of the options presented by the algorithm were appropriate. The wizard indicated that he did not often choose to disregard the option of sending utterances to ongoing conversation, even when the topic of conversation was different than presented by the system. Analysis of system logs reveals that the wizard chose to remain silent less than 1% of the time when prompted by the algorithm. The wizard noted that he felt the need to be involved during conversation and not remain “silent” for more than a few turns, and thus chose one of the options presented by the system, even if he found them inadequate. We asked the wizard what he would do differently in such situations, and he noted that simply switching to a new theme or continuing to talk about the previous theme made it seem that he (as a participant) was deliberately ignoring the theme being discussed by the other participants, and also made it seem that he deliberately not being involved in that discussion.

*Finding 4:* The main critique given by the wizard during debriefings was that the agent had no options to offer in situations where “detrimental” topics were raised by other participants.

5.5.5 *Conclusion from WOZ experiments*

Qualitative responses indicated that the wizard was perceived to be persuasive by a majority of participants and was also ranked highly in the peer-ratings of persuasion. There were a sizable proportion (44%) of instances when the agent strategies were successful in leading to attitude change, as characterized by numerical changes in self-ratings of belief strength and belief evaluation. However, the resultant attitude change could not be captured quantitatively as the difference score was found to be statistically non-significant.
in our sample of subjects. The most likely interpretation is that the algorithm did not anticipate the presence of detrimental topics, which led to a negative persuasion effect in the participant’s belief models.

Our Findings 1 through 4 allowed us to infer one important fact from the WOZ experiments: **well-defined counter-persuasion mechanisms need to be explicitly encoded in the algorithm in order for the system to be considerably more successful.**

Subsequently, we engaged in algorithm refinements and conducted our next series of experiments. In these experiments, the agent acted autonomously. We describe the refinements and experiment results next.

### 5.6 Data Analysis and Results - Autonomous agent experiments

In this section, we will present the data analysis from the autonomous agent experiments. First, in Section 5.6.1, we elaborate upon the changes made to the agent algorithm based on the results from our WOZ experiments.

#### 5.6.1 Refining the algorithm

We discussed in the Literature Review chapter (Chapter 2, Section 2.4.3) that Topic Control behavior showed the highest correlation with the sociolinguistic phenomena of Influence. Additionally, introduction of new topics that are subsequently discussed by others during conversation was found to be the strongest indicator of topic control behavior. Consequently, our revised counter-persuasion strategies capitalize upon these two pivotal discoveries. We made a straightforward refinement to the algorithm: if a detrimental topic is introduced in conversation, the agent deliberately steers the conversation away to a new topic. Doing so potentially attributes a greater degree of topic control behavior to the agent and detracts from the topic control attributed to other participants.
In order to motivate this refinement, we draw on Schank’s (1977) *Rules and Topics in Conversation* and Schank and Abelson’s work (Schank and Abelson, 2013) where purposefully shifting the conversation to new topics is a strategy for persuasion. The rules for topic shift given by Schank, to a large extent, are no more than rules for the first step in free association of ideas. The METATOPIC, as characterized by Schank: “A metatopic is a comment that can be inferred from the interaction of two conceptualizations.” can be used to shift the topic, to a greater degree of dissociation from the previous topic of conversation, than a topic whose conceptualization shares some common subset with the previous topic. In our counter-persuasion strategy, we make use of the METATOPIC to change the topic of conversation, in that, we will attempt to steer the conversation to a new theme, which had not been discussed before and is divorced from the previous topic. The algorithmic implementation of counter-persuasion strategy is described next.

We added two new behaviors to the Behavior Selector module (Chapter 4, Section 4.5). In Table 5.6.1, we describe the triggers associated with eliciting such behaviors in the agent. The two new behaviors are *Counteract Detrimental Theme* and *Make Assertive Statement on New Theme*. These two behaviors are executed in a sequence of moves by the agent, in response to the introduction of a detrimental theme by any participant in the conversation. Any *Assertion-Opinion*, *Information-Request* or *Confirmation-Request* made by any participant on a detrimental theme will trigger this behavior in the agent. The belief evaluation of such utterances would not be aligned with the agent model. If the agent detects that a pre-designated “detrimental” topic/theme has been introduced during conversation, it will engage in a series of moves for counteracting the detrimental topic.
<table>
<thead>
<tr>
<th>Dialogue Act</th>
<th>Communication Link</th>
<th>$b_i$</th>
<th>$e_i$</th>
<th>Theme</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assertion-Opinion;</td>
<td>any</td>
<td>any</td>
<td>not aligned with</td>
<td>detrimental</td>
<td>Counteract Detrimental Theme; Make Assertive Statement on New Theme</td>
</tr>
<tr>
<td>Information-Request;</td>
<td></td>
<td></td>
<td>agent model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confirmation-Request</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.6.1. Behaviors made by the agent in order to counteract persuasion

**Counter-Persuasion Behavior**

When this behavior is induced, the agent will make two utterances in the conversation, in a sequence. The *Counteract Detrimental Theme* has the following templates.

**Templates to choose from:**

2. *<person X>, I don’t think that is very likely to be true*

3. *I think that is very unlikely, <person X>*

4. *<person X>, I don’t think that is quite likely*

5. *I believe that is quite unlikely, <person X>*

6. *<person X>, that is very unlikely*

The *Counteract Detrimental Theme* behavior is intended to steer the conversation away, in spite of not offering any arguments to do so.

An alternative counter-persuasion strategy would be to extend the list of arguments and themes as shown in Table 5.6.2 and include arguments for the detrimental themes as well, so that an exhaustive list of arguments related to the topic can be enumerated. We could obtain such arguments by conducting a targeted search on the web for text related to the topic. The retrieved results could be content-analyzed following procedures similar to those outlined in the belief annotation procedures in Chapter 3. We designate such an alternative counter-persuasion strategy as a potential avenue for future research and
further elaborate on it in the next chapter of this thesis (Chapter 6, Conclusions and Future Work). In designing our autonomous agent (AA) experiments, we wanted to be as faithful as possible to the experiment design used during WOZ experiment. The primary reason was to ensure the comparability of experimental results across both conditions, and that relevant overarching conclusions could be drawn.

Following a statement using the Counteract Detrimental Theme behavior, the agent will attempt to steer the conversation to a new topic by engaging in the Make Assertive Statement on New Theme behavior next using one of the templates below.

**Templates to choose from:**

1. Consider this instead - <text>; 2. Instead, I feel that <text>; 3. Actually, if you think about it, <text>; 4. Instead, don’t you think <text>; 5. Instead, we should consider <text>

The Make Assertive Statement on New Theme is similar to Make Assertive Statement behavior, which is already part of the agent’s repertoire, but it attempts to explicitly steer the conversation in a new direction. The agent also simulates a “typing” delay between the two consecutive utterances to appear plausibly human (the typing speed is simulated to be 1 word per second). The new theme is selected in the pre-determined order in which the themes are imported by the system prior to the chat session. As discussed in Section 4.6 (Chapter 4), the themes are listed in order of most common belief targets based on participant pre-discussion survey responses and the number of arguments available for the theme.

After incorporating these refinements in the agent algorithm, we conducted our next series of experiments, the AA experiments. In these experiments, there was no human wizard.
The agent acted autonomously during the chat sessions. There were other changes to the experimental procedure from the WOZ chat sessions. 12 participants completed the AA chat sessions. The following sections describe the results related to these experiments, starting with descriptive statistics.

5.6.2 Descriptive Statistics

We conducted four chat sessions in the Autonomous Agent (AA) condition. These are named agent_session_1 through agent_session_4. We present descriptive statistics on these sessions in Table 5.6.2.1. The topic of discussion and stance of the agent on the topic was determined by the group composition and the belief models of the participants.

The total number of utterances in these chat sessions (Mean=194.25, stdev=22.60) was higher than in the WOZ experiments (Mean=177.83, stdev=23.91). Additionally, the number of utterances made by the agent in the AA condition (Mean=38.25, stdev=6.07) versus in the WOZ condition (Mean=29.5, stdev=5.82) was also found to be higher. Importantly, there were no statistically significant differences (p=0.30 for the total number of utterances in chat sessions, and p=0.87 for number of utterances made by the agent) across the AA and WOZ conditions. The higher number of utterances made by the agent could be attributed to the longer length of discussions. Figure 5.6.2.1 shows the turn distribution of all participants across all AA chat sessions. We find statistically significant differences in the number of turns made by river and the agent zen (p<=0.05). The differences in the number of turns made by star and memphis were not statistically significant when compared to the number of turns made by zen (p<=0.05). The average rank of zen on the sociolinguistic behavior of involvement computed using the DSARMD
system was 2.25. Thus, the agent was not necessarily uninvolved in the conversation, even though it was making fewer turns than the other participants.

<table>
<thead>
<tr>
<th>Discussion Topic</th>
<th>agent_session_1</th>
<th>agent_session_2</th>
<th>agent_session_3</th>
<th>agent_session_4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy access to morning after pill to people under 17</td>
<td>41 (26%)</td>
<td>45 (29%)</td>
<td>36 (23%)</td>
<td>31 (20%)</td>
</tr>
<tr>
<td>Lowering minimum legal drinking age to 18 from 21</td>
<td>64 (49%)</td>
<td>63 (49%)</td>
<td>42 (42%)</td>
<td>54 (45%)</td>
</tr>
<tr>
<td>Total utterances across all participants</td>
<td>213</td>
<td>214</td>
<td>170</td>
<td>180</td>
</tr>
</tbody>
</table>

Table 5.6.2.1 Overview of the four chat sessions conducted during Autonomous Agent experiments

In Figure 5.6.2.2, we show the distribution of utterances made by the agent across all AA chat sessions. The utterance distribution is similar to the distribution observed during
WOZ experiments. We also observe that the agent made five instances of the counter-persuasion behavior (which is a sequence of two moves), shown in the last column of Figure 5.6.2.2. There were no statistically significant differences found in the distribution of utterances across both experiments, WOZ and AA (p=0.69).

Finding 1: There were no statistically significant differences in the total number of utterances in a chat session; the total utterances produced by the agent; and the types of utterances across the AA and WOZ conditions. This finding is not surprising, given the design of our system and the experiment. It serves as a validation that the system behaves as expected when acting in autonomous mode.
5.6.3 Perception Ratings

On the perception question, which specifically asked participants whether they thought *zen* was persuasive, we found that a majority of the participants rated *zen* to be persuasive. This data, presented in Figure 5.6.3.1 shows that eight (67%) participants either agreed or strongly agreed with the statement ‘Participant *zen* was persuasive’. Three participants either disagreed or strongly disagreed with the statement ‘Participant *zen* was persuasive’ (25% of the sample population).

![Distribution of responses on question: Participant *zen* was persuasive](image)

**Figure 5.6.3.1** Participant responses on whether *zen* was persuasive or not

For the peer ratings of persuasiveness, we found that the average rank given to *zen* to be 2.25, across all sessions (1 being the most persuasive, 4 being the least). Automated rankings on influence from DSARMD system (Strzalkowski et al., 2010; 2012) gave *zen* an average rank of 1.25. A strong positive correlation (Pearson’s *R*=0.80) was observed
between automated system rankings and peer-rankings on the persuasiveness of zen (although it should be noted here that the sample n=4 chat sessions is small).

**Finding 2:** A substantial proportion of participants perceived zen to be persuasive (66.7%). The average rank given to the agent in AA condition (2.25) on persuasiveness was slightly higher than that given to the agent in the wizard condition (2.33). A strong positive correlation is observed between automated rankings and peer-ranking on persuasiveness. This finding is consistent with Finding 2 from the WOZ experiments. We also compared the peer-ratings given to the system in WOZ experiments to those in the AA experiments, and found the results to be statistically significant (t=3.28, p=0.005).

### 5.6.4 Persuasion Effectiveness

We observed statistically significant changes\(^5\) in overall attitude from pre- and post-discussion surveys using the Wilcoxon signed rank test (Wilcoxon \(T = 12\), \(p = 0.03\)). We note here that the results should be interpreted in context of the relatively small sample size of n=12. If the sample size is at least 20, then the Wilcoxon test statistic \(W\) tends to form a normal distribution, in which case the \(Z\)-value can be used to evaluate the results. Otherwise, the \(W\)-value is used. We present both the \(Z\)-value and \(W\)-value in the results shown in the footnote below. A complete explanation of the Wilcoxon test formula and calculations on the data from the WOZ and AA experiments are presented in Appendix E.

In Figure 5.6.4.1, we show the absolute change in attitudes of participants as the difference of \(\sum b_i e_i\) from pre- and post-discussion surveys. On the x-axis are the 12 participants who

\(^5\) Result 1 - Z-value
The Z-value is -2.1181. The p-value is 0.034. The result is significant at \(p \leq 0.05\).

Result 2 - W-value
The W-value is 12. The critical value of \(W\) for \(N = 12\) at \(p \leq 0.05\) is 13. Therefore, the result is significant at \(p \leq 0.05\).
participated in the AA experiments. The y-axis denotes their change in attitude (average=15.167, stdev=19.144). A majority of participants changed in favor of the position argued by the agent.

![Absolute attitude change seen in participants across all WOZ sessions](image)

**Figure 5.6.4.1 Change in attitudes for all twelve AA session participants**

![AA Session 1: Easy access to Morning After Pill to people under 17](image) ![AA Session 2: Easy access to Morning After Pill to people under 17](image)

**Figure 5.6.4.2 and Figure 5.6.4.3 Change in attitudes for participants in AA session 1 and session 2**

In Figures 5.6.4.2-5.6.4.5, we show the changes in pre- to post-discussion attitudes for all participants in each AA chat session. We see that the agent is successful in persuading all the participants in sessions 1 and 4. In session 2, we see that the attitude of participant
river shifts slightly in favor of the agent, while star and memphis move away from the agent. In session 3, while the agent is successful in persuading river and memphis, participant star moves away from the agent.

Figure 5.6.4.4 and Figure 5.6.4.5 Change in attitudes for participants in AA session 3 and session 4

Note that the key difference between the agent algorithm in the AA condition and the WOZ condition was the addition of the counter-persuasion strategy. Thus, of primary interest is the study of instances when the agent engaged in counter-persuasive behavior during the discussions. We illustrate the counter-persuasion behaviors exhibited by the system via specific examples below.

In Figure 5.6.4.6, we show a brief exchange between river and zen on the topic of MAP: Easy access to morning after pill to people under 17 taken from agent_session_1. Participant river introduces the theme of Teenagers may abuse MAP. The sequence of moves executed by the agent is shown next. The agent makes an assertion on the theme MAP allows teenagers more control over their sex life. For this conversation, this new theme was the next in sequence in the list of pre-ordered themes imported by the agent at the outset of
conversation. In this instance, the theme *Teenagers may abuse MAP* may be considered (at least partially) circumvented, because the discussion no longer centers around that topic.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Pre-discussion Response</th>
<th>Post-discussion Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>river</em></td>
<td>bi</td>
<td>ei</td>
</tr>
<tr>
<td>river pre-discussion response</td>
<td>3</td>
<td>-3</td>
</tr>
<tr>
<td>river post-discussion response</td>
<td>2</td>
<td>-3</td>
</tr>
<tr>
<td><em>star</em></td>
<td>bi</td>
<td>ei</td>
</tr>
<tr>
<td>star pre-discussion response</td>
<td>2</td>
<td>-3</td>
</tr>
<tr>
<td>star post-discussion response</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td><em>memphis</em></td>
<td>bi</td>
<td>ei</td>
</tr>
<tr>
<td>memphis pre-discussion response</td>
<td>1</td>
<td>-3</td>
</tr>
<tr>
<td>memphis post-discussion response</td>
<td>0</td>
<td>-3</td>
</tr>
</tbody>
</table>

Figure 5.6.4.6 Fragment of conversation showing agent’s counter-persuasion moves

In Table 5.6.4.1, we show the change in bi*ei for participants on the theme *Teenagers may abuse MAP* from the same discussion as presented in the example above. Participant *river*
changed her belief strength to a small extent (pre-discussion belief strength +3, very likely to +2, likely post-discussion). Participant star changed her belief strength from +2 (likely) to 0 (undecided). The net effect of this change is that such detrimental theme does not factor into star’s overall attitude towards the topic any longer (bi*ei=0). A similar condition is observed for participant memphis, although their belief strength in the pre-discussion survey was weak (+1, somewhat likely).

In Figure 5.6.4.7, we show another instance of counter-persuasion strategy used by the agent on the same theme in a different chat session, agent_session_2. In this example, participant river continues to discuss the detrimental theme, even after the counter-persuasion strategy is invoked by the agent. This is an example of unsuccessful counter-persuasion because the agent is unable to deter the conversation occurring on the detrimental theme.

<table>
<thead>
<tr>
<th>river (10:08:10 am): But I think this also increases the probability of its abuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>THEME: Teenagers may abuse the MAP</td>
</tr>
<tr>
<td>zen (10:08:30 am): river, that is very unlikely.</td>
</tr>
<tr>
<td>zen (10:08:44 am): Consider this instead - it is safe and also recommended by the World Health Organization</td>
</tr>
<tr>
<td>THEME: MAPs are safe</td>
</tr>
<tr>
<td>river (10:09:02): It is recommended by them but abuse is not a good thing.</td>
</tr>
<tr>
<td>THEME: Teenagers may abuse the MAP</td>
</tr>
<tr>
<td>memphis (10:09:47): yeah i am sure there have not been enough studies on the pill and it could cause potential harm to the woman’s body</td>
</tr>
<tr>
<td>THEME: Teenagers may abuse the MAP</td>
</tr>
</tbody>
</table>

Figure 5.6.4.7 Fragment of conversation showing agent’s counter-persuasion moves and subsequent discussion
In this conversation, the new theme chosen by the agent is *MAPs are safe*. This is a new theme chosen by the agent according to the order in which the themes were pre-programmed at the outset of conversation. In the previous example, the agent was able to successfully counter-persuade. We can infer that the theme chosen in the previous example was more effective in counter-persuading than the theme in example shown in Figure 5.6.4.7. In the current implementation of the system, we do not know *apriori* the themes that will be more effective when presenting counter-persuasive arguments and the selection of the counter-persuasive theme can therefore by considered somewhat arbitrary. As part of future work, we expect that the agent can make more informed choices of themes specifically for the counter-persuasion behaviors.

<table>
<thead>
<tr>
<th>Teenagers may abuse MAP</th>
<th>bi</th>
<th>ei</th>
<th>bi*ei</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>river</em> pre-discussion response</td>
<td>2</td>
<td>-3</td>
<td>-6</td>
</tr>
<tr>
<td><em>river</em> post-discussion response</td>
<td>3</td>
<td>-3</td>
<td>-9</td>
</tr>
<tr>
<td><em>star</em> pre-discussion response</td>
<td>2</td>
<td>-2</td>
<td>-4</td>
</tr>
<tr>
<td><em>star</em> post-discussion response</td>
<td>3</td>
<td>-3</td>
<td>-9</td>
</tr>
<tr>
<td><em>memphis</em> pre-discussion response</td>
<td>1</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td><em>memphis</em> post-discussion response</td>
<td>2</td>
<td>-2</td>
<td>-4</td>
</tr>
</tbody>
</table>

Table 5.6.4.2 Result of unsuccessful counter-persuasion by the agent on a detrimental theme

The resultant changes in bi*ei* for the participants in this session for the theme of *Teenagers may abuse MAP* are shown below in Table 5.6.4.2. All participants move away from the intended agent model on this particular theme, suggesting that the agent counter-persuasion strategy failed.
On the other hand, the trends of successful persuasion that we had observed during the WOZ experiments held true in the AA experiments as well. These trends where the agent could successfully persuade participants to change belief strength (when belief evaluation was aligned), and change belief evaluation (when belief strength was weak, have been discussed in Section 5.5.3. This was reassuring, because we did not want to impair or detract from the prior successful strategies of the algorithm observed during the WOZ experiments when we added the counter-persuasion refinement.

**Finding 3:** Statistically significant changes in attitudes were observed during the AA experiments. Although, the results from this experiment should be conservatively interpreted due to lack of sufficient sample size (twelve participants across four chat sessions), this finding is encouraging. The only difference in the underlying algorithm in the AA experiments was the addition of the counter-persuasion strategy. Thus, we posit that the increased success of the agent in these experiments can be attributed to the refinement we added, although further, careful experimentation will be required to sufficiently validate this claim.

### 5.6.5 Qualitative interviews with participants

We conducted qualitative post-discussion interviews with the participants. Almost all participants indicated that they found the experience enjoyable (one participant rated the experience Neither enjoyable nor unenjoyable). Across all chat sessions, the agent’s legitimacy as a human was not compromised.

Below in Figure 5.6.5.1, we present a few notable excerpts from the interviews. With respect to feedback on experiment design, one participant noted that the time given for discussion should be increased and another noted that the discussion might be facilitated
or structured somewhat. Two participants noted that, in hindsight, they could perceive oddities in participant *zen*'s behavior, but only after they were made aware of the presence of an artificial agent by the experimenter during debriefing. In future experiments, we wish to experiment with increasing the time allowed for discussion. One additional behavior in the agent's repertoire in the future might be to facilitate or structure the discussion, acting as a task leader (Broadwell et al., 2013) and giving it an additional measure of authoritative influence.

|“I knew right away that *zen* was the participant not in this room” |
|“could I do it one more time? I want to see if I can persuade the bot in my favor” |
|“*zen* seemed odd to me, and now I know why.” |
|“Longer time duration should be given to conclude properly.” |
|“Maybe there could be a sub-topic or specific questions to help facilitate the discussion.” |
|“I thought *zen* was a guy!” |
|“Did not realize there was a robot in the chat room” |

Table 5.6.5.1 Excerpts from qualitative interviews with participants

**Finding 4:** Participants enjoyed the experience and were not aware that they were conversing with an artificial agent until they were made aware during the debriefing.

**5.7 Summary of this chapter**

In this chapter we presented the details of validating our approach. Two series of experiments were conducted. Following the WOZ experiments, we executed a
straightforward, yet significant refinement of adding explicit counter-persuasion strategies to the system. The results from the AA experiments showed statistically significant attitude change in participants. We present our conclusions from this thesis in Chapter 6.
Chapter 6: Conclusion

6.1 Introduction

We set out to answer two research questions (RQs) in this thesis. Restated from Chapter 1, the questions are:

1. Can specific persuasive strategies be automated in a virtual chat agent?
2. Can active persuasion by individuals during conversation be detected and counteracted by such an agent?

In order to tackle RQ1, we designed a virtual chat agent, equipping it with three major modules (Chapter 4). We grounded the persuasion strategies in well-accepted social influence theories, and programmed these at the very core of the agent algorithm (Chapter 2). We empirically validated its efficacy under controlled conditions (Chapter 3 and Chapter 5). We also incorporated counter-persuasion strategies into the algorithm (Chapter 5) following analyses from initial experiments using a human wizard. Our autonomous agent experiments have shown statistically significant attitude change resulting from successful persuasion attributed to the agent.

6.2 Future Directions

Several avenues for future research are viable. These are listed below:

6.2.1 Further Validation against larger sample size

The sample size in our experiments was relatively small. If we want to establish true empirical validity of our results, the sample size required would be approximately 384 participants\(^6\). We could accomplish this in two ways – we could either follow the same

\(^6\) Calculated using parameters \(\alpha=5\%\); Confidence Level=95\% at: http://www.raosoft.com/samplesize.html
experiment design we have discussed in Chapter 5, with three participants and one agent in each chat session, and conduct a total of 128 chat sessions. Alternatively, we could increase the number of participants per session, to 6 participants and conduct 64 chat sessions (4 majority and 2 minority). Whilst the time and resources allowed for this thesis do not enable us to conduct experiments on these scales, we wish to test our conclusions on the recommended sample size as part of future research.

6.2.2 Further validation using a control condition
An additional consideration towards establishing empirical validity is including a control condition as part of our experimental design. Building upon the suggested research avenue in 6.2.1, we could use a 2 X 2 factorial design, by varying group size and the presence of the agent in the group as our dependent variables.
In effect, our null hypothesis in this research has been that the interventions made by the agent have no effect on participant attitudes. We have assumed that if the agent were not present during discussion, the participant responses on pre-discussion and post-discussion survey would not change. However, we need to verify this hypothesis by conducting experiments where such a control condition is explicitly made part of the experiment design.

6.2.3 Inoculation Experiments
Another compelling future avenue for research is employing multiple persuasion agents, working in concert to collaboratively achieve certain tasks with conversation. For example, a secondary persuasion agent could act as an ally of the primary, and support the statements made by the primary agent, thereby reinforcing the effects of persuasion on the other participants. One other way to design the persuasive strategies for the multiple agents is to draw upon research in inoculation theories. McGuire (1964) suggested that attitudes could
be inoculated against persuasive attacks in much the same way that a person’s immune system can be inoculated against viral attacks. Inoculation messages involve two primary components that foster attitudinal resistance among recipients: threat and refutational preemption (McGuire, 1964). Threat refers to a potentially negative message towards the attitude (Compton, 2013). The second feature of an inoculation treatment is refutational preemption. This component of a message “provides specific content that receivers can employ to strengthen attitudes against subsequent change” in response to the threat noted above (Compton, 2013) In most inoculation messages, refutational preemption is characterized by the raising and refuting of counterarguments (i.e., challenges to an existing position). Thus, a conventional inoculation message begins with a forewarning of impending challenges to a held position (threat), then raises and refutes some possible challenges that might be raised by opponents. In our experiments, we could deploy two agents, one whose task would be to generate messages perceived as threats and another whose task would be to refute those messages by offering counterarguments.

6.2.4 Validation “in the wild”

While our controlled experiments are well suited to provide us with evidence of the efficacy of the agent, real-world experiments under unconstrained settings are called for to truly test the agent’s capabilities. For instance, we could deploy the persuasion agent on social media platform, such as Twitter, where its goal would be to persuade followers. Such experiments would need to be carefully designed, as the ethical implications are considerably complicated, and would also require redesigning the algorithm to manage asynchronous communication over a public channel using 140 characters or less. This future research goal,

---

7 www.twitter.com
while challenging, is extremely compelling, because it would establish the validity of the agent beyond controlled laboratory experimental settings.

6.3 Contributions of this thesis

The research questions we posed have not been systematically studied in prior research, in the manner described in this dissertation. In addition, the experiments to implement and validate the research in this thesis are discernibly novel. It is our contention that the work described in this thesis favorably impacts the fields of computational modeling of human behavior, artificial intelligence, human-computer interaction and cognition. The data and software resulting from this thesis is freely available for download.\(^8\)

\(^8\) www.samirashaikh.com/
References:


Harlow, S. (2013). It was a" Facebook revolution": Exploring the meme-like spread of narratives during the Egyptian protests. Revista de comunicación,12.


Appendix A: Copy of Belief Elicitation Survey

Demographic Information
Please provide some basic demographic information about yourself.

First Name: *

Last Name: *

Email (this will the address you will receive correspondence on): *

Age: *

Gender *

Field of Study/Intended Major:
(Optional)

« Back   Continue »
What do you think are the DISADVANTAGES of genetically modified food? *
Please provide as elaborate a response as you can.

What do you think are the ADVANTAGES of genetically modified food? *
Please provide as elaborate a response as you can.

What do you think are the ADVANTAGES of stricter airport screening procedures? *
Please provide as elaborate a response as you can.

What do you think are the DISADVANTAGES of stricter airport screening procedures? *
Please provide as elaborate a response as you can.
Appendix B: Copy of pre-discussion survey

**CONSENT FORM GOES HERE**

Please insert your initials in the box below. By doing so, you confirm that you have read the above information and consent to participate in this research. Please feel free to download and keep a copy for your records.*

1) First Name:* __________________________

2) Last Name:* __________________________

3) Email:* __________________________

4) Gender:*
() Female
() Male
() Prefer not to answer

5) Age:* __________________________

6) Intended Major: __________________________

7) In order for us to understand more about identity in online discussion groups, please tell us what you consider to be your ethnicity. Select all that apply: *
() American Indian
() Arab, Egyptian or Maghreb
() Black/African
() East Asian - e.g. Chinese, Japanese, South-East Asian
() Hispanic/Latino
() South Asian - e.g. Indian, Pakistani, Bangladeshi
() Pacific Islander
() White/Caucasian/European
() Other (please specify): __________________________
Topic: Genetically Modified Foods

8) How likely do you feel the following statement to be true?

*Genetically modified foods have a harmful effect on health:*
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

9) *Having a harmful effect on health is:*
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

10) How likely do you feel the following statement to be true?

*Genetically modified foods reduce starvation:*
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

11) *Reducing starvation is:*
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

12) How likely do you feel the following statement to be true?

*Genetically modified foods lead to more affordable food:*
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

13) *More affordable food is:*
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

14) How likely do you feel the following statement to be true?

*Genetically modified foods cause greater quantities of food to be produced:*
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

15) *Producing greater quantities of food is:*
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good
16) How likely do you feel the following statement to be true?

*Genetically modified foods have a harmful effect on environment:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

17) A harmful effect on environment is:

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

18) How likely do you feel the following statement to be true?

*Genetically modified foods are adaptable to many climates:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

19) Being adaptable to many climates is:

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

20) How likely do you feel the following statement to be true?

*Genetically modified foods are detrimental to farmers:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

21) Being detrimental to farmers is:

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

22) How likely do you feel the following statement to be true?

*Genetically modified foods have not been properly researched:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

23) Not being properly researched is:

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

24) How likely do you feel the following statement to be true?

*Genetically modified foods are healthier than non-GM foods:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely
25) **Being healthier than non-GM foods**

   ( ) Very Bad   ( ) Bad   ( ) Poor   ( ) Neither Good nor Bad   ( ) Fair   ( )
   Good   ( ) Very Good

26) **How likely do you feel the following statement to be true?**

   **Genetically modified foods taste better than non-GM foods:**

   ( ) Very unlikely   ( ) Unlikely   ( ) Somewhat unlikely   ( ) Undecided   ( ) Somewhat likely   ( ) Likely   ( ) Very likely

27) **Tasting better than non-GM foods**

   ( ) Very Bad   ( ) Bad   ( ) Poor   ( ) Neither Good nor Bad   ( ) Fair   ( )
   Good   ( ) Very Good

28) **How likely do you feel the following statement to be true?**

   **Genetically modified foods are fake or unnatural:**

   ( ) Very unlikely   ( ) Unlikely   ( ) Somewhat unlikely   ( ) Undecided   ( ) Somewhat likely   ( ) Likely   ( ) Very likely

29) **Being fake or unnatural**

   ( ) Very Bad   ( ) Bad   ( ) Poor   ( ) Neither Good nor Bad   ( ) Fair   ( )
   Good   ( ) Very Good

30) **How likely do you feel the following statement to be true?**

   **Genetically modified foods have fewer chances of bacteria or disease than non-GM foods:**

   ( ) Very unlikely   ( ) Unlikely   ( ) Somewhat unlikely   ( ) Undecided   ( ) Somewhat likely   ( ) Likely   ( ) Very likely

31) **Having fewer chances of bacteria or disease than non-GM foods**

   ( ) Very Bad   ( ) Bad   ( ) Poor   ( ) Neither Good nor Bad   ( ) Fair   ( )
   Good   ( ) Very Good

How likely would you be to vote in favor of more genetically modified foods?*

   ( ) Very unlikely   ( ) Unlikely   ( ) Somewhat unlikely   ( ) Undecided   ( ) Somewhat likely   ( ) Likely   ( ) Very likely
Topic: Stricter airport screening procedures

32) How likely do you feel the following statement to be true?

Stricter airport screening procedures increase safety:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

33) Increasing safety is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

34) How likely do you feel the following statement to be true?

Stricter airport screening procedures increases risk of racial or gender profiling:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

35) Increasing risk of racial or gender profiling is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

36) How likely do you feel the following statement to be true?

Stricter airport screening procedures increase time spent at airports:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

37) Increasing time spent at airports is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

38) How likely do you feel the following statement to be true?

Stricter airport screening procedures increase invasion of privacy:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

39) Increasing invasion of privacy is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

40) How likely do you feel the following statement to be true?

Stricter airport screening procedures increase peace of mind for travelers:*

191
41) Increasing peace of mind for travelers is:
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

42) How likely do you feel the following statement to be true?

Strict airport screening procedures reduces chances of terrorism:
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

43) Reducing chances of terrorism is:
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

44) How likely do you feel the following statement to be true?

Strict airport screening procedures increases chances of finding weapons:
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

45) Increasing chances of finding weapons is:
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

46) How likely do you feel the following statement to be true?

Strict airport screening procedures deter people from traveling in airplanes:
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

47) Deterring people from traveling in airplanes is:
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

48) How likely do you feel the following statement to be true?

Strict airport screening procedures increase costs to travelers:
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

49) Increasing costs to travelers is:
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good
50) How likely do you feel the following statement to be true?

_Stricter airport screening procedures increase employment for the TSA:*_

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

51) _Increasing employment for the TSA is:*_

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

How likely would you be to vote in favor of stricter airport screening procedures?*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely
Topic: Easy access to morning after pill for people under 17

52) How likely do you feel the following statement to be true?

*Easy access to morning-after pill for people under 17 leads to fewer teenage pregnancies:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

53) *Fewer teenage pregnancies are:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

54) How likely do you feel the following statement to be true?

*Easy access to morning-after pill for people under 17 will lead to more unprotected sex amongst teenagers:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

55) *Unprotected sex amongst teenagers is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

56) How likely do you feel the following statement to be true?

*Easy access to morning-after pill for people under 17 leads to increased sexual activity in teenagers:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

57) *Increased sexual activity in teenagers is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

58) How likely do you feel the following statement to be true?

*Teenagers may abuse the morning after pill:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

59) *Teenagers abusing the morning-after pill is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good
60) How likely do you feel the following statement to be true?

*Easy access to morning-after pill for people under 17 lead to increased risk of sexually transmitted diseases:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

61) *Increased risk of sexually transmitted diseases is:*

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

62) How likely do you feel the following statement to be true?

*Easy access to morning-after pill for people under 17 lead to adverse effects on menstrual cycle:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

63) *Adverse effects on menstrual cycles is:*

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

64) How likely do you feel the following statement to be true?

*Easy access to morning-after pill for people under 17 allows teenagers more control over their sex life:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

65) *Allowing control over their sex life for teenagers is:*

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

66) How likely do you feel the following statement to be true?

*Easy access to morning-after pill for people under 17 avoids parental involvement in decision to use the pill:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

67) *Avoiding parental involvement in decision to use morning-after pill is:*

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good
68) How likely do you feel the following statement to be true?

*Easy access to morning-after pill for people under 17 are last resort safety nets for teenagers:*

( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

69) *Last resort safety nets for teenagers is:*

( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

70) How likely do you feel the following statement to be true?

*Easy access to morning-after pill for people under 17 leads to fewer abortions:*

( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

71) *Fewer abortions are:*

( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

How likely would you be to vote in favor of easy access to morning after pill for people under 17?*

( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely
Topic: Lowering minimum legal drinking age from 21 to 18

72) How likely do you feel the following statement to be true?

*Lowering the minimum legal drinking age would reduce alcohol abuse:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

73) *Reducing alcohol abuse is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

74) How likely do you feel the following statement to be true?

*Lowering the minimum legal drinking age would increase drunk driving incidents:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

75) *Drunk driving incidents are:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

76) How likely do you feel the following statement to be true?

*Lowering the minimum legal drinking age leads to acknowledgment of 18 as age of adulthood:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

77) *Acknowledging 18 as the age of adulthood is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

78) How likely do you feel the following statement to be true?

*Lowering the minimum legal drinking age would lower rate of crimes related to alcohol:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

79) *Lowering rate of crimes related to alcohol is:*
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good
80) How likely do you feel the following statement to be true?

*Lowering the minimum legal drinking age would increase risk of alcoholism:*
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

81) *Increasing risk of alcoholism is:*
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

82) How likely do you feel the following statement to be true?

*Lowering the minimum legal drinking age leads to more money in the economy:*
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

83) *More money in the economy is:*
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

84) How likely do you feel the following statement to be true?

*Lowering the minimum legal drinking age would have adverse effects on developing brains:*
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

85) *Adverse effects on developing brains are:*
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

86) How likely do you feel the following statement to be true?

*Lowering the minimum legal drinking age is a slippery slope (potentially making even younger kids able to drink)::*
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

87) *Being a slippery slope (potentially making even younger kids able to drink) is:*
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

88) How likely do you feel the following statement to be true?

*Lowering the minimum legal drinking age would decrease underage drinking:*
89) **Decreasing underage drinking is:**

- ( ) Very Bad
- ( ) Bad
- ( ) Poor
- ( ) Neither Good nor Bad
- ( ) Fair
- ( ) Good
- ( ) Very Good

90) **How likely do you feel the following statement to be true?**

*Lowering the minimum legal drinking age increases risk of irresponsible behavior:*

- ( ) Very unlikely
- ( ) Unlikely
- ( ) Somewhat unlikely
- ( ) Undecided
- ( ) Somewhat likely
- ( ) Likely
- ( ) Very likely

91) **Increasing risk of irresponsible behavior is:**

- ( ) Very Bad
- ( ) Bad
- ( ) Poor
- ( ) Neither Good nor Bad
- ( ) Fair
- ( ) Good
- ( ) Very Good

92) **How likely do you feel the following statement to be true?**

*Lowering the minimum legal drinking age increases risk of sexual assaults:*

- ( ) Very unlikely
- ( ) Unlikely
- ( ) Somewhat unlikely
- ( ) Undecided
- ( ) Somewhat likely
- ( ) Likely
- ( ) Very likely

93) **Increasing risk of sexual assault is:**

- ( ) Very Bad
- ( ) Bad
- ( ) Poor
- ( ) Neither Good nor Bad
- ( ) Fair
- ( ) Good
- ( ) Very Good

94) **How likely do you feel the following statement to be true?**

*Lowering the minimum legal drinking age increases rate of college dropouts:*

- ( ) Very unlikely
- ( ) Unlikely
- ( ) Somewhat unlikely
- ( ) Undecided
- ( ) Somewhat likely
- ( ) Likely
- ( ) Very likely

95) **Increasing rate of college dropout is:**

- ( ) Very Bad
- ( ) Bad
- ( ) Poor
- ( ) Neither Good nor Bad
- ( ) Fair
- ( ) Good
- ( ) Very Good

**How likely would you be to vote in favor of lowering the minimum legal drinking age from 21 to 18?**

- ( ) Very unlikely
- ( ) Unlikely
- ( ) Somewhat unlikely
- ( ) Undecided
- ( ) Somewhat likely
- ( ) Likely
- ( ) Very likely
Topic: Living together before marriage

96) How likely do you feel the following statement to be true?

*Living together before marriage leads to knowing partner at a deeper level:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

97) Knowing partner at a deeper level is:
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

98) How likely do you feel the following statement to be true?

*Living together before marriage leads to knowing if there is compatibility:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

99) Knowing if there is compatibility is:
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

100) How likely do you feel the following statement to be true?

*Living together before marriage leads to sharing responsibility and finances:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

101) Sharing responsibility and finances is:
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

102) How likely do you feel the following statement to be true?

*Living together before marriage increases chances of breakup:*
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

103) Increasing chances of breakup is:
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good
104) How likely do you feel the following statement to be true?

*Living together before marriage increases chances of partners getting tired of one another:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

105) *Increasing chances of partners getting tired of one another is:*

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

106) How likely do you feel the following statement to be true?

*Living together before marriage increases knowledge of partner's habits:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

107) *Increasing knowledge of partner's habits is:*

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

108) How likely do you feel the following statement to be true?

*Living together before marriage is a test-run before married life:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

109) *A test-run before married life is:*

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

110) How likely do you feel the following statement to be true?

*Living together before marriage is disapproved by society or religion:*

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

111) *Being disapproved by society or religion is:*

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

112) How likely do you feel the following statement to be true?

*Living together before marriage avoids need for divorce:*

201
113) **Avoiding need for divorce is:**
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

114) How likely do you feel the following statement to be true?

*Living together before marriage reduces the excitement of marriage:*
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

115) **Reducing the excitement of marriage is:**
( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

**How likely would you be to vote in favor of living together before marriage?**
( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely
Topic: Being able to easily identify who you are interacting with online

116) How likely do you feel the following statement to be true?

**Being able to easily identify who you are interacting with online reduces freedom of expression:**

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

117) Reducing freedom of expression is:

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

118) How likely do you feel the following statement to be true?

**Being able to easily identify who you are interacting with online decreases privacy:**

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

119) Decreasing privacy is:

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

120) How likely do you feel the following statement to be true?

**Being able to easily identify who you are interacting with online reduces cyber-bullying:**

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

121) Reducing cyber-bullying is:

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good

122) How likely do you feel the following statement to be true?

**Being able to easily identify who you are interacting with online leads to creation of fake persona online:**

( ) Very unlikely  ( ) Unlikely  ( ) Somewhat unlikely  ( ) Undecided  ( ) Somewhat likely  ( ) Likely  ( ) Very likely

123) Creating a fake persona online is:

( ) Very Bad  ( ) Bad  ( ) Poor  ( ) Neither Good nor Bad  ( ) Fair  ( ) Good  ( ) Very Good
124) How likely do you feel the following statement to be true?

*Being able to easily identify who you are interacting with online holds people accountable for their actions:*

( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

125) *Holding people accountable for their actions is:*

( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

126) How likely do you feel the following statement to be true?

*Being able to easily identify who you are interacting with online improves interaction and community building:*

( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

127) *Improving interaction and community building is:*

( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

128) How likely do you feel the following statement to be true?

*Being able to easily identify who you are interacting with online decreases information security:*

( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

129) *Decreasing information security is:*

( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

130) How likely do you feel the following statement to be true?

*Being able to easily identify who you are interacting with online leads to more honesty online:*

( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

131) *More honesty online is:*

( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good
132) How likely do you feel the following statement to be true?

Being able to easily identify who you are interacting with online reduces trolling activity:*  
( ) Very unlikely × ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

133) Reducing trolling activity is:*  
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

134) How likely do you feel the following statement to be true?

Being able to easily identify who you are interacting with online reduces spread of false information and rumors:*  
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely

135) Reducing spread of false information and rumors is:*  
( ) Very Bad ( ) Bad ( ) Poor ( ) Neither Good nor Bad ( ) Fair ( ) Good ( ) Very Good

How likely would you be to vote in favor of being able to easily identify who you are interacting with online?*  
( ) Very unlikely ( ) Unlikely ( ) Somewhat unlikely ( ) Undecided ( ) Somewhat likely ( ) Likely ( ) Very likely
Topic Information

136) How informed are you about the following topics? Please rank each topic in order (1=most informed, 6= least informed). You can only choose one response per topic:*
   ______ Living together before marriage
   ______ Genetically modified foods
   ______ Easier access to morning-after pill to people under 17
   ______ Being able to easily identify who you are interacting with online
   ______ Stricter airport screening procedures
   ______ Lowering minimum legal drinking age from 21 to 18

Topic Summary

137) How interested are you in discussing the following topics? Please rank each topic in order of your level of interest (1=most interested, 6= least interested). You can only choose one response per topic:*
   ______ Living together before marriage
   ______ Genetically modified foods
   ______ Easier access to morning-after pill to people under 17
   ______ Being able to easily identify who you are interacting with online
   ______ Stricter airport screening procedures
   ______ Lowering minimum legal drinking age from 21 to 18

138) Please indicate your availability for participating in a chat session from the options below.
We will schedule a session when at least 4 people have signed up for the same time slot.

PLEASE SELECT AS MANY SLOTS AS POSSIBLE. *
[ ] Monday, April 4th 2016 *10 am - 11:30 am*
[ ] Monday, April 4th 2016 *11:30 am - 1 pm*
[ ] Tuesday, April 5th 2016 *12:00 pm - 1:30 pm*
[ ] Tuesday, April 5th 2016 *1:30 pm - 3 pm*
[ ] Wednesday, April 6th 2016 *10 am - 11:30 am*
[ ] Wednesday, April 6th 2016 *11:30 am - 1 pm*
[ ] Wednesday, April 6th 2016 *1 pm - 2:30 pm*
[ ] Thursday, April 7th 2016 *12 noon - 1:30 pm*
[ ] Thursday, April 7th 2016 *1:30 pm - 3 pm*
[ ] Friday, April 8th 2016 *10 am - 11:30 am*
[ ] Friday, April 8th 2016 *11:30 am - 1 pm*

Thank You!
Appendix C: Copy of post-discussion survey

1) First Name:*  
_________________________________________________

2) Last Name:*  
_________________________________________________

3) Email:*  
_________________________________________________

4) Your id during the chat session was:*  

( ) Zen  
( ) River  
( ) Star  
( ) Memphis
25) During the discussion, some people are more persuasive than others. Below is a list of participants, including yourself.

Please rate each participant in terms of how persuasive they seemed to you:
* 

_____ Memphis
_____ Zen
_____ Star
_____ River

26) One participant from the discussion is selected at random for this question. Please indicate your opinion about this participant.

Participant Zen was persuasive.*

( ) Strongly Agree   ( ) Agree   ( ) Neither Agree nor Disagree   ( ) Disagree   ( ) Strongly Disagree

27) How enjoyable did you find this experience?*

( ) Very Enjoyable   ( ) Enjoyable   ( ) Neither Enjoyable nor Unenjoyable   ( ) Unenjoyable   ( ) Very Unenjoyable

28) Please provide any comments or feedback for us to improve this experience in the future (optional):

_________________________________________________

Thank You!
Appendix D: Debriefing Form

Debriefing Form for Participation in Research Study
ARCSS: Advanced Research in Computational Social Science

Thank you for your participation in our study! Your participation is greatly appreciated.

Purpose of the Study:

Earlier in our consent form we informed you that the purpose of the study was to collect samples of realistic on-line chat room discussions on a number of topics related to current issues in education, technology, arts, sports, finances, politics, etc. In actuality, our study is about testing automated agents in a multi-party chat setting and seeing if the chat agents are undetected and can achieve certain goals in conversation.

Unfortunately, in order to properly test our hypothesis, we could not provide you with all of these details prior to your participation. This ensures that your reactions in this study were spontaneous and not influenced by prior knowledge about the purpose of the study. If we had told you the actual purposes of our study, your ability to respond to the conversation and react to the agent naturally could have been affected. We regret the deception but we hope you understand the reason for it.

Confidentiality:

Please note that although the purpose of this study is different from the originally stated purpose, everything else on the consent form is correct. This includes the ways in which we will keep your data confidential. Your name will not be associated with any of the information that you provide during the course of this study. You will be assigned a unique and anonymous identification code during this study and your demographic information will be kept separate from that. All data collected will be maintained on a secure server, ensuring that your identity is completely safe.

Now that you know the true purpose of our study and are fully informed, you may decide that you do not want your data used in this research. If you would like your data removed from the study and permanently deleted please let us know and we will securely delete all of your data.

Whether you agree or do not agree to have your data used for this study, you will still receive the intended credit for your participation.

Please do not disclose research procedures and/or hypotheses to anyone who might participate in this study in the future as this could affect the results of the study.
Final Report:

If you would like to receive a copy of the final report of this study (or a summary of the findings) when it is completed, please feel free to contact us.

If you have any further questions or concerns regarding this study, its purpose or procedures, or if you have a research-related problem, please feel free to contact the researchers at tomek@albany.edu or sshaikh@albany.edu.

***Please keep a copy of this form for your future reference. Once again, thank you for your participation in this study!***
Appendix E: Wilcoxon signed rank test calculations

Formula and steps⁹:

Steps in Obtaining the Wilcoxon Signed-Ranks Test Statistic \( W \)

1. For each item in a sample of \( n \) items obtain a difference score \( D_i \) between two measurements.
2. Neglect the “+” and “−” signs and obtain a set of \( n \) absolute differences \( |D_i| \).
3. Omit from further analysis any absolute difference score of zero, thereby yielding a set of \( n' \) nonzero absolute difference scores, where \( n' \leq n \). Thus, \( n' \) becomes the actual sample size—after we have removed observations with absolute difference scores of zero.
4. Then assign ranks \( R_i \) from 1 to \( n' \) to each of the \( |D_i| \) such that the smallest absolute difference score gets rank 1 and the largest gets rank \( n' \). Owing to a lack of precision in the measuring process, if two or more \( |D_i| \) are equal, they are each assigned the average rank of the ranks they would have been assigned individually had ties in the data not occurred.
5. Now reassign the symbol “+” or “−” to each of the \( n' \) ranks \( R_i \), depending on whether \( D_i \) was originally positive or negative.
6. The Wilcoxon test statistic \( W \) is obtained as the sum of the positive ranks [see Equation (10.11)].

WILCOXON SIGNED-RANKS TEST STATISTIC \( W \)

The Wilcoxon test statistic \( W \) is obtained as the sum of the positive ranks.

\[
W = \sum_{i=1}^{n'} R_i^{(+)} \tag{10.11}
\]

⁹ Taken from: http://courses.wcupa.edu/rbove/Berenson/CD-ROM%20Topics/topice-10_5.pdf
Calculations for WOZ experiment results

<table>
<thead>
<tr>
<th>Pre-discussion attitude</th>
<th>Post-discussion attitude</th>
<th>Sign</th>
<th>Abs</th>
<th>R</th>
<th>Sign R</th>
</tr>
</thead>
<tbody>
<tr>
<td>-36</td>
<td>5</td>
<td>-1</td>
<td>41</td>
<td>17</td>
<td>-17</td>
</tr>
<tr>
<td>-3</td>
<td>22</td>
<td>-1</td>
<td>25</td>
<td>16</td>
<td>-16</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>-1</td>
<td>20</td>
<td>14</td>
<td>-14</td>
</tr>
<tr>
<td>-11</td>
<td>6</td>
<td>-1</td>
<td>17</td>
<td>13</td>
<td>-13</td>
</tr>
<tr>
<td>18</td>
<td>29</td>
<td>-1</td>
<td>11</td>
<td>10</td>
<td>-10</td>
</tr>
<tr>
<td>-5</td>
<td>0</td>
<td>-1</td>
<td>5</td>
<td>5</td>
<td>-5</td>
</tr>
<tr>
<td>17</td>
<td>19</td>
<td>-1</td>
<td>2</td>
<td>2.5</td>
<td>-2.5</td>
</tr>
<tr>
<td>-18</td>
<td>-17</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>-31</td>
<td>-35</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>-22</td>
<td>-29</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>-3</td>
<td>-11</td>
<td>1</td>
<td>8</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>-4</td>
<td>-12</td>
<td>1</td>
<td>8</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>69</td>
<td>60</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>-12</td>
<td>-27</td>
<td>1</td>
<td>15</td>
<td>11.5</td>
<td>11.5</td>
</tr>
<tr>
<td>2</td>
<td>-13</td>
<td>1</td>
<td>15</td>
<td>11.5</td>
<td>11.5</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>1</td>
<td>24</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>18</td>
<td>-38</td>
<td>1</td>
<td>56</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

Result Details

*W*-value: 78.5
Mean Difference: -20.56
Sum of pos. ranks: 92.5
Sum of neg. ranks: 78.5

*Z*-value: -0.3049
Mean (W): 85.5
Standard Deviation (W): 22.96
Sample Size (N): 18

*Result 1 - Z-value*
The *Z*-value is -0.3049. The p-value is 0.76418. The result is not significant at p ≤ 0.05.

*Result 2 - W-value*
The *W*-value is 78.5. The critical value of *W* for N = 18 at p ≤ 0.05 is 40. Therefore, the result is not significant at p ≤ 0.05.
Calculations for AA experiment results

<table>
<thead>
<tr>
<th>Pre-discussion attitude</th>
<th>Post-discussion attitude</th>
<th>Sign</th>
<th>Abs</th>
<th>R</th>
<th>Sign R</th>
</tr>
</thead>
<tbody>
<tr>
<td>-26</td>
<td>12</td>
<td>-1</td>
<td>38</td>
<td>12</td>
<td>-12</td>
</tr>
<tr>
<td>-14</td>
<td>22</td>
<td>-1</td>
<td>36</td>
<td>11</td>
<td>-11</td>
</tr>
<tr>
<td>-4</td>
<td>28</td>
<td>-1</td>
<td>32</td>
<td>10</td>
<td>-10</td>
</tr>
<tr>
<td>-33</td>
<td>-4</td>
<td>-1</td>
<td>29</td>
<td>9</td>
<td>-9</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>-1</td>
<td>28</td>
<td>7.5</td>
<td>-7.5</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>-1</td>
<td>28</td>
<td>7.5</td>
<td>-7.5</td>
</tr>
<tr>
<td>-31</td>
<td>-14</td>
<td>-1</td>
<td>17</td>
<td>6</td>
<td>-6</td>
</tr>
<tr>
<td>17</td>
<td>19</td>
<td>-1</td>
<td>2</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>-18</td>
<td>-17</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>-3</td>
<td>-11</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>-13</td>
<td>1</td>
<td>15</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Result Details

W-value: 12
Mean Difference: -29.58
Sum of pos. ranks: 12
Sum of neg. ranks: 66

Z-value: -2.1181
Mean (W): 39
Standard Deviation (W): 12.75
Sample Size (N): 12

Result 1 - Z-value
The Z-value is -2.1181. The p-value is 0.034. The result is significant at p≤ 0.05.

Result 2 - W-value
The W-value is 12. The critical value of W for N = 12 at p≤ 0.05 is 13. Therefore, the result is significant at p≤ 0.05.
Appendix F: Dialogue act tags

A. Statements and Responses

A Statement makes a claim about the world, and tries to change the beliefs of the listener. In general, an utterance that is a statement can be said to be true or false. Responses are utterances that users make to indicate reaction to another user’s utterances, such as answering it, acknowledging it or agreeing to it.

We partition the Statements and Responses into four different types of dialogue acts – Assertion-Opinion, Offer-Commit, Acknowledge and Signal-non-understanding.

A.1 Assertion-Opinion

Assertion-Opinions are statements that communicate some specific details or make a claim about the world.

Example 27.

User 1. (1) This is my first attendance.

(DA: Assertion-Opinion)

Example 28.

User 2. (2) I chat pretty regularly, lots of long-distance friends

(DA: Assertion-Opinion)

Example 29.

User 1. (1) sure….I was not totally bummed out by the Watchmen movie like I thought I would be....

User 1. (2). I kind of thought I would hate it, because I'm so attached to the book.

(CL: Continuation-of:User1:1, DA: Assertion-Opinion)

Example 30.

User 3. (3) its more effective, I think

(DA: Assertion-Opinion)

The above examples illustrate the various forms an utterance with an Assertion-Opinion tag can take. Example 27, 28 and 29 are general statements providing some information to the other users while in Example 30, the user is expressing an opinion.
When we take into account the context in which a user is making an assertion-opinion type of utterance, it may be categorized as providing an answer to a question that another user has asked, or they may be agreeing with an opinion stated by another user. In such cases we use one of the specialized tags under the Assertion-Opinion label, if any of them applies: Response-Answer, Response-Non-Answer, Agree-Accept or Disagree-Reject.

A.1.1 Response-Answer

A response to a question asked by another user, that fulfills the question is tagged as a Response-Answer. In the examples below, User 2’s utterance is tagged as Response-Answer. In general, a Response-Answer tag is applied to an utterance that is a Response-to a prior utterance.

Example 31.
User 1. (1) wait, you have to go at what time?
User 2. (2) like 1240 or so….
(CL: RESPONSE-TO-User1:1, DA: RESPONSE-ANSWER)

Example 32.
User 1. (1) Who created this money? Is it backed by a local bank or something?
User 2. (2) It’s backed by some kind of local money…
(CL: RESPONSE-TO-User1:1, DA: RESPONSE-ANSWER)

Example 33.
User 3 (1) do you guys have an Albany Craigslist?
User 4. (2) yes sir! love it!
(CL: RESPONSE-TO-User3:1, DA: RESPONSE-ANSWER)

In Examples 34 and 35, although User 2’s utterance is structured as a reply in the negative, it ‘answers’ the question asked by User 1 and it is tagged as a Response-Answer.

Example 34.
User 1. (1) …you on youtube right now?
User 2. (2) no but if you paste a link ill check out whatever
(CL: RESPONSE-TO-User1:1, DA: RESPONSE-ANSWER)

Example 35.
User 3. (3) is there some kind of occassion today?
User 4. (4) no every sunday we do brunch
(CL: RESPONSE-TO-User3:3, DA: RESPONSE-ANSWER)
A general rule to follow is if an utterance is identified as being a ‘response-to’ another utterance that was a question, and it satisfies the question that was asked, we select this tag.

**A.1.2 Response-Non-Answer**

A response to a question asked by another user, that does not fulfill the question is tagged as a Response-Non-Answer. This occurs when the requested answer is not supplied, either because it is unknown, or irrelevant, or does not apply, or when it is withheld. In general, a Response-Non-Answer tag will be applied to an utterance that is a Response-to a prior utterance.

In Examples 36 and 37, both utterances by User 2 are tagged as Response-Non-Answer:

**Example 36.**

User 1. (1) what song?

User 2. (2) I don’t remember, it was something with Lil’ Wayne on it

(CL: RESPONSE-TO-User1:1, DA: RESPONSE-NON-ANSWER)

**Example 37.**

User 3. (1) is there a real italian place in albany? buccas maybe?

User 4. (2) umm not sure i am lead to believe there is because there are alot of italians there haha

(CL: RESPONSE-TO-User3:1, DA: RESPONSE-NON-ANSWER)

In example 38 below, User 2’s response-to User 1’s utterance is tagged as a Response-Non-Answer, because it does not directly answer the question posed. We note that a simple ‘No’ response would be labeled as Response-Answer.

**Example 38.**

User 1. (5) did anyone watch the morning talk shows today (MTP, for example)?

User 2. (6) i don’t have cable :

(CL: RESPONSE-TO-User1:5, DA: RESPONSE-NON-ANSWER)

For assigning Response-Non-Answer tag, a general rule to follow is: if an utterance is identified as being a ‘response-to’ another utterance that was a question and it does not satisfy the question that was asked, we select this tag.
A.1.3 Agree-Accept

We mark an utterance with this tag if the user accepts or agrees with another user’s proposal or request; or if the information or claim conveyed in a statement is accepted or confirmed. Statements with this tag convey the meaning - ‘I agree with you’, ‘I will accept what you say to be true’ or ‘I will go along with what you propose’

Example 39.
User 1. (1) saw is overflowing with cruelty!
(DA:ASSERTION-OPINION)
User 2. (2) for sure.
(CL: RESPONSE-TO-User1:1, DA: AGREE-ACTCEPT)

Example 40.
Lance. (1) I guess in real life I might just ring her up and clear up any possible misunderstandings before I nix her.
Jennifer. (2) absolutely lance.
(CL: RESPONSE-TO-LANCE:1, DA: AGREE-ACTCEPT)

Example 41.
User 1. (1) yeah richard is better than Emily
User 2. (2) I think so too.
(DA: AGREE-ACTCEPT)

Example 42.
User 1. (1) that’s a nice thing to do but might not be as relevant to the job description as some of the other stuff
User 2. (2) yeah relevancy to the description is #1
(DA: AGREE-ACTCEPT)

The utterances in the above examples made by User 2 are examples of Agree-Accept utterances.

A.1.4 Disagree-Reject

We use this tag when the user disagrees, rejects a proposal or offer, says he will not comply, or says that the claim or the information expressed by the other user is incorrect. Statements with this tag convey the meaning - ‘I disagree with you’, ‘I do not accept what you say to be true’ or ‘I will not go along with what you propose’.

217
Example 43.
   amanda. (3) Kara, try just walking around. I love NY, to coin a phrase.
   kara. (4) no i can barely take the city life of albany to be honest
   (CL: RESPONSE-TO-AMANDA:3, DA: DISAGREE-REJECT)

Example 44.
   User 1. (1) Maybe the ILS Personnel person will come out of hiding!
   User 2. (2) i don’t think so
   (DA: DISAGREE-REJECT)

Example 45.
   User 1. (1) Benny’s seriously subtle!
   User 2. (2) i guess there’s subtle aspects, but not when he’s slapping the old bald
dude on the head!
   (CL: RESPONSE-TO-User1:1, DA: DISAGREE-REJECT)

Example 46.
   User 3. (1) Sushi tastes great!
   User 4. (2) only when dipped in soy sauce though.
   (CL: RESPONSE-TO-User3:1, DA: DISAGREE-REJECT)

Example 47.
   User 3. (3) internet does not correct your spelling mistakes, ....
   User 4. (4) not unless you use a program
   (CL: RESPONSE-TO-User3:3, DA: DISAGREE-REJECT)

The utterances in the above examples made by User 2 are examples of Disagree-Reject utterances.

A.2 Offer-Commit
We assign Offer-Commit tags to utterances that may potentially obligate the speaker to a future action. Offers are implicit or explicit questions that, if answered in the affirmative or with some positive information, mean that the speaker will perform some action for the listener. The Commits are utterances in which the speaker obligates himself/herself to perform a future action, phrased in such a way that the commitment is not contingent on the listener’s agreement. A Commit may also be the response to an Action-Directive (dialogue act b.3).
Example 48.
User 1. (1) Also, I think I will review the resume once again . . .
(DA: OFFER-COMMIT)

Example 49.
User 3. (3) So I’ll broaden my horizons and listen.
(DA: OFFER-COMMIT)

Example 50.
User 1. (1) I’ll give it a shot.
(DA: OFFER-COMMIT)

Example 51.
User 1. (1) ok i’ll look up the charity.
(DA: OFFER-COMMIT)

Example 52.
User 1. (1) I’ll go look at Edwards.
(DA: OFFER-COMMIT)

Examples 48 and 49 are types of offers. Examples 50 – 52 are types of commit. We assign all these utterances the Offer-Commit tag.

A.3 Acknowledge
An Acknowledge tag is applied to utterances that recognize, without necessarily accepting or rejecting a previous utterance made by another user. A short phrase such as “okay”, “yes” or “uh-huh” can indicate that the user heard and understood the previous utterance, but did not necessarily accept what he heard.

Example 53.
User 1. (1) goods we buy have to increase in price since our money is worth less
User 2. (2) I see.
(CL: RESPONSE-TO-User1:1, DA: ACKNOWLEDGE)

Example 54.
User 4. (2) oh nice.
(CL: RESPONSE-TO-User3:1, DA: ACKNOWLEDGE)
Example 55.
User 1. (1) get this...there is a cafe in the saratoga region
User 2. (2) yes?
(DA: ACKNOWLEDGE)

Example 56.
User 1. (1) That was important in prior chats when the screen for inputting characters was limited . .
User 2. (2) mmhmm
(DA: ACKNOWLEDGE)

Example 57.
User 1. (1) i lived in japan for 7 years
User 2. (2) cool!
(DA: ACKNOWLEDGE)

In all of the above examples, User 2’s utterances are assigned the Acknowledge tag. We also include in this category utterances such as ‘lol’, ‘haha’, ‘funny!’, ‘weird!’ that also convey user’s reception of a prior utterance.

A.4 Signal-non-understanding
If the user has not understood or has partially understood something, we use the Signal-non-understanding tag. Utterances of this type can usually be paraphrased as “I don’t understand”, or “What did you mean?” or “Is this what you said?” or “I beg your pardon?” Many of these utterances take form of questions but they should not be confused with Information-Request or Confirmation-Request dialogue acts.

Example 58.
User 1. (1) it is related to affective aspects in an interactive design product, such as chat
User 2. (2) Huh? You lost me Jenny. Affective aspects?
(CL: RESPONSE-TO-User1:1, DA: SIGNAL NON-UNDERSTANDING)

Example 59.
User 1. (1) Yes, I liked Osama (the movie!) better, but Kite was good.
User 2. (2) osama the movie? :o
(DA: SIGNAL NON-UNDERSTANDING)
Example 60.

User 1. (1) he’s an mma superstar from quebec

User 2. (2) MMA?

(CL: RESPONSE-TO-User1:1, DA: SIGNAL-NON-UNDERSTANDING)

User 1. (3) mixed martial arts, ultimate fighting championship

(CL: RESPONSE-TO-User2:2, DA: RESPONSE-ANSWER)

In all of the above examples, User 2’s utterances are assigned the Signal-non-understanding tag. Unlike an information request (e.g., ‘where in Japan?’ or ‘do you know Hulk Hogan?’), this dialogue act does not introduce or asks for new information; instead it often repeats a phrase mentioned in a prior utterance to which it responds. The use of discourse particles such as “huh?” or emoticons such as “:o” is common as well. Signal-non-understanding may sometimes appear similar to Confirmation-Request (next section); annotators should exercise their best judgment as to whether the speaker is asking for clarification or merely wants to reconfirm a fact. In the example below User2 asks for confirmation, which is then answered by User1. This may be contrasted with Example 60.1 where User2 is confirming whether the meeting starts at 7 pm or not. Asking the question ‘MMA?’ means that the user has not understood what is meant by MMA as opposed to the question ‘7?’ where they are attempting to confirm time when the meeting will start.

Example 60.1: User 1. (1) the meeting will start at 7 tomorrow

User2. (2) 7?

(DA: CONFIRMATION-REQUEST)

User1. (3) Yup

(DA: RESPONSE-ANSWER)

B. Questions and Directives

This dimension characterizes what effect an utterance has on the subsequent dialogue and interaction. Questions and directives are typically those that elicit some response from other users.

We classify questions and directives into the following four categories – Information-Request, Confirmation-Request, Open-Question-Option and Action-Directive.
B.1 Information-Request
An utterance where the user is asking for new information is tagged as Information-Request. These are typically questions that require a response with information that is new or previously unknown to the asker. In the examples below, the users are asking for information they do not currently possess. All these examples are tagged as Information-Request. We also include in this category questions that are more general in nature such as asking for opinions on an item or issue such as in examples 66 and 67 below.

Example 61.
User 1. (1) where is everyone?
(CL: ADDRESSED-TO-ALL-USERS, DA: INFORMATION-REQUEST)

Example 62.
User 2. (2) who is leading tonight?

Example 63.
User 3. (3) who is Gov of NY now?

Example 64.
User 4. (4) How can you browse and do this?

Example 65.
User 5. (5) what’s a leader’s responsibilities?

Example 66.
User 6. (6) what about text messaging
(DA: INFORMATION-REQUEST)

Example 67.
User 7. (7) what about cell phone use in general
(CL: ADDRESSED-TO-ALL-USERS, DA: INFORMATION-REQUEST)

B.2 Confirmation-Request
A Confirmation-Request is an utterance that calls for the listener to confirm (or to deny) a fact that is already known to the speaker (Examples 68, 69). Confirmation-request can be used to verify whether an assumption made by the speaker is correct or not (Example 69). A Confirmation-Request typically takes the form of a non-inverted yes/no question and is often accompanied by a tag question such as “right?” or “didn’t you?”, etc. Confirmation-Request should not be confused with those yes-no questions in which the speaker requests some new information of the listener (User 4’s utterance in Example 70). Such questions should be tagged as Information-Request. Annotators should base their judgment
upon which facts appear to be known to each dialogue participant based on the immediate context (usually 2-3 utterances prior).

Example 68.
User 1. (1) You running the show today Nick?
(CL: ADDRESSED-TO-NICK, DA: CONFIRMATION-REQUEST)

Example 69.
User 2. (2) sarah, you just visited Seattle?
(DA: CONFIRMATION-REQUEST)

Example 70.
User 3. (3) How much do you travel?
(DA: INFORMATION-REQUEST)
User 4. (4) Does back and forth to work count?
(CL: RESPONSE-TO-User3:3, DA: INFORMATION-REQUEST)

Example 71.
User 5. (5) its your husband who doesn’t like it, right?
(DA:CONFIRMATION-REQUEST)

Example 72.
User 6. (6) have you seen old boy?
(DA: INFORMATION-REQUEST)

B.3 Action-Directive
If the user directs another user to perform some action, we label the utterance as Action-Directive.

Example 73.
User 1.(1) ok leave! and comeback
(DA: ACTION-DIRECTIVE)

Example 74.
User 2. (2) well everyone wish me luck on my midterm
(DA: ACTION-DIRECTIVE)

Example 75.
User 3.(3) Take a break first!
(DA: ACTION-DIRECTIVE)

Example 76.
User 1. (1) Let’s move on to Richard’s resume now.
(DA: ACTION-DIRECTIVE)
Example 77.
User 2. (2) you should try to find some riders on there, it would make gas super cheap, and it makes the drive pretty interesting.
(DA: ACTION-DIRECTIVE)

Example 78.
User 3. (3) yeah you should ask them about him, amanda, haha
(DA: ACTION-DIRECTIVE)

An Action-Directive places an obligation upon the listener to respond: either to perform the action as requested or to refuse it. An Action-Directive would often be followed by either Agree-Accept or Disagree-Reject utterances.

C. Conversational Norms

This group of tags deals with the social norms occurring in conversations. We place Conventional-Opening, Conventional-Closing, Thanking and Apology in this category. Selecting a tag from this category automatically selects the Communication-Management Information-Level tag.

C.1 Conventional-Opening

The Conventional-Opening tag indicates that the user is beginning the interaction by using a conventional social phrase to greet another user, or by replying to such a greeting with a conventional phrase.

Example 79.
User 1. (1) Hello Jenny. How are you?
(CL: ADDRESSED-TO-JENNY, DA: CONVENTIONAL-OPENING)

Example 80.
User 2. (2) Hi Alex!
(CL: ADDRESSED-TO-ALEX, DA: CONVENTIONAL-OPENING)

Example 81.
User 3. (3) good evening

Example 82.
User 4. (4) Hi all!
C.2 Conventional-Closing
The Conventional-Closing label is used for utterances in which the user utters a conventional social phrase or expression to finish or wrap up the conversation as seen in the examples below.

Example 83.
User 1. (1) i’m headed out as well; nice talking to you all!
(DA: CONVENTIONAL-CLOSING)

Example 84.
User 2. (2) ok bye bye
(DA: CONVENTIONAL-CLOSING)

Example 85.
User 3. ok guys was fun .... good night!
(DA: CONVENTIONAL-CLOSING)

C.3 Other-Conventional-Phrase
Social phrases used in conversation that do not fall under the Conventional-Opening and Conventional-Closing categories are assigned this label. These are utterances that are considered appropriate responses in a conventional social setting such as apologies, praise, thanking etc (User 2’s utterance in Examples 86, 87 and 88 below).

Example 86.
User 1. (1) Good luck with your midterm!
User 2. (2) THANKS!!!!!!
(CL: RESPONSE-TO-User1:1, DA: OTHER-CONVENTIONAL-PHRASE)

Example 87.
User 1. (1) I am not feeling well at all today.
User 2. (2) I am so sorry to hear that.
(CL: RESPONSE-TO-User1:1, DA: OTHER-CONVENTIONAL-PHRASE)

Example 88.
User 1. (1) Hi, how are you?
User 2. (2) Fine, thanks!
(CL: RESPONSE-TO-User1:1, DA: OTHER-CONVENTIONAL-PHRASE)

In contrast, however, consider Example 89 below. Although the utterance ‘thanks, alex!’ uses the words ‘thanks’ that may be an indicator of a conventional phrase, it actually conveys Kara’s acceptance of Alex’s offer to look up a website and it is coded as an Agree-Accept.
Example 89.

alex. (1) I will look up the website for you.
kara. (2) thanks, alex!

(CL: RESPONSE-TO-ALEX, DA: AGREE-ACCEPT)

C.4 Correct-Misspelling

A Correct-Misspelling tag will be applied to utterances that are entered to correct a misspelling in a previous utterance.

Example 90.

User 1. (1) Yeah i’ve heard of that phenomenan becfore
User 1. (2) before*

(CL: CONTINUATION-OF-User1:1, DA: CORRECT-MISSPELLING,
IL: COMMUNICATION-MANAGEMENT)

Example 91.

User 2. (1) OK let’s have htem both in for interviews
User 1. (2) *them!

(CL: CONTINUATION-OF-User2:1, DA: CORRECT-MISSPELLING,
IL: COMMUNICATION-MANAGEMENT)

We see in example 91 that the user misspelled the word “them” in the first utterance and the second utterance typed the correct spelling. The second utterance in example 90 and 91 will be labeled ‘correct-misspelling’. Usually such corrections are preceded or followed by an asterisk (*), though this may not always be the case.