Quantitative Framework For Social Cultural Interactions

2016

Taranjeet Singh Bhatia
University of Central Florida

Quantitative Framework For Social Cultural Interactions

by

TARANJEET SINGH BHATIA
B.E. Instrumentation Engineering, Rajiv Gandhi Technical University, Bhopal 2007
M.S. Computer Science, University of Central Florida, Orlando 2013

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2016

Major Professor: Ladislau Bölöni

ABSTRACT

For an autonomous robot or software agent to participate in the social life of humans, it must have a way to perform a calculus of social behavior. Such a calculus must have explanatory power (it must provide a coherent theory for why the humans act the way they do), and predictive power (it must provide some plausible events from the predicted future actions of the humans).

This dissertation describes a series of contributions that would allow agents observing or interacting with humans to perform a calculus of social behavior taking into account cultural conventions and socially acceptable behavior models. We discuss the formal components of the model: culture-sanctioned social metrics (CSSMs), concrete beliefs (CBs) and action impact functions. Through a detailed case study of a crooked seller who relies on the manipulation of public perception, we show that the model explains how the exploitation of social conventions allows the seller to finalize transactions, despite the fact that the clients know that they are being cheated. In a separate study, we show that how the crooked seller can find an optimal strategy with the use of reinforcement learning.

We extend the CSSM model for modeling the propagation of public perception across multiple social interactions. We model the evolution of the public perception both over a single interaction and during a series of interactions over an extended period of time. An

important aspect for modeling the public perception is its propagation - how the propagation is affected by the spatio-temporal context of the interaction and how does the short-term and long-term memory of humans affect the overall public perception.

We validated the CSSM model through a user study in which participants cognizant with the modeled culture had to evaluate the impact on the social values. The scenarios used in the experiments modeled emotionally charged social situations in a cross-cultural setting and with the presence of a robot. The scenarios model conflicts of cross-cultural communication as well as ethical, social and financial choices. This study allowed us to study whether people sharing the same culture evaluate CSSMs at the same way (the inter-cultural uniformity conjecture). By presenting a wide range of possible metrics, the study also allowed us to determine whether any given metric can be considered a CSSM in a given culture or not.

Ik Oankaar Sathigur Prasaadh

One Universal Creator God. By The Grace Of The True Guru:

This work is dedicated to the most wonderful parents in the world, Surjeet Singh and Rajinder Kaur for their constant love, patience, and care. I can't begin to thank you for all you have done over the years for me. Your sacrifices, support and prayers made it possible.

ACKNOWLEDGMENTS

I am most grateful to the members of my committee, Dr. Damla Turgut, Dr. Gita Suk-thankar, and Dr. Steve Fiore for their time, encouragement, and expertise throughout this work.

Special thanks to the chairman of my committee, Dr. Ladislau Bölöni, for his exquisite attention to detail and for his demand for excellence. He has been an amazing mentor and I'm extremely grateful for his ongoing feedback, support and consistent push for critical thinking that challenged me throughout this work.

There are people in everyone's lives who make success both possible and rewarding. A special thanks and feeling of gratitude towards my loving sisters, Manpreet Kaur and Satbeer Kaur who have never left my side and are very special.

TABLE OF CONTENTS

ix

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Achieving appropriate behavior in a social-cultural context is one of the most elusive goals of agent research. There are, however, many practical applications where social behavior is necessary. Agents acting in virtual environments, such as games or training must show a believable social behavior. This can often be achieved with careful scripting. However, when agents control autonomous robots which interact with humans in social settings, the interactions are more open ended and rigid scripting is not possible. Furthermore, the agent or the robot does not only need to act in a believable way, but it must actually convey meaning and achieve goals through social interaction. Thus, the agent must have a model through which it can evaluate the impact of specific actions on the participants in the social interaction. There are actions that are physically possible, but socially unacceptable in a given culture. We will use the term *social calculus* for this evaluation process.

To be useful in such scenarios, the social calculus must be expressed in an operational, algorithmic form, suitable for software implementation. Although the literature of social sciences contains many sophisticated models of social and cultural behavior, these models rely heavily on the judgement of a human observer. Software implementations do not benefit from the experience of the human observer, thus the models we are looking for must be more algorithmic and less reliant on interpretative nuances.

A further requirement is that features should be acquired in a non-invasive way from sensing (vision and voice recognition). At the current stage of the technology we can rely on the robots to identify the locations of humans in the scene, recognize simple gestures, find some key-words in spoken text and identify the loudness and tone of voice. We can not, however, assume that a system can match the capacity of humans for identifying facial expressions, and perform deep semantic processing of the spoken communication.

The objective of this dissertation is to develop social calculus techniques that can be implemented by an agent with the currently available sensing technology and reasoning models. Although the ability to model human reasoning with pure mathematics (the famous "calculemus" of Leibniz [1]) is still out of reach, we found that a careful mixture of formal models and knowledge engineering can yield practically deployable models with explanatory and predictive power.

The work in this dissertation extends an earlier model of culture and concrete belief framework [2]. In Chapter 3, we describe a formal model for the representation of social interactions [3, 4, 5]. We assume that the human behavior proceeds through a series of actions a_i. Actions impact the state of the actor, the target of the action, their peers as well as the perception of the general public. In this model, the state of the agent, relevant to its actions in the social-cultural context is described by a collection of *metrics*. The metrics can be divided into *tangibles* (such as wealth and time) and *socially constructed* (such as dignity, politeness, generosity or kindness). CSSMs are not necessarily independent, but they are not arbitrarily convertible to each other as discussed in Section 3.2. CSSMs provide a relatively

high-detail model of the social behavior: in its spirit, this technique falls close to the KIDS (Keep it Descriptive Stupid) approach advocated by Edmonds and Moss [6].

The impact of an action on a CSSM is not a constant. Rather, it is modulated by the beliefs of the agent about specific aspects of the current context. A culture requires its members to maintain these beliefs as accurate as possible - the correctness of beliefs is necessary for the culture to operate as expected. Nevertheless, it is quite possible for an agent to have incorrect beliefs, especially in inter-cultural exchanges, when the agent might misinterpret the social signals (computers are especially bad at this, see ([7])). As agents will act and calculate CSSMs according to the beliefs, we need to trace the belief values even when they are not correct. If an agent considers another one as friend, it will act accordingly and judge the actions of the other agent in this context, regardless whether the friendship is mutual or not. We discussed these beliefs in Section 3.3.

A critical component of this model is the set of action-impact functions (AIFs) that describe how the actions of the agents change the CSSMs in specific settings. AIFs are multi-parameter mathematical functions, we can not directly ask them from human informants. Knowledge engineering these functions for every possible action is a difficult challenge, because the design space is very large. The Spanish Step scenario as described in Section 3.1 has only two participants - yet there are 20 different actions and 14 different CSSMs (if we consider self, peer and public perceptions separately). This is already a significant knowledge engineering task. As we are moving to more open-ended scenarios, with a larger number of participants, the number of AIFs and their respective complexity increases at least quadrat-

ically. Finding efficient methods to acquire the AIFs is thus a critical step in making the CSSM approach applicable to medium size real world interaction scenarios. We elaborated on this in Section 3.4.

CSSMs are consistent in a given culture, but they vary between cultures. A given culture assigns a name, a calculation method and a series of behavior rules to these metrics. Agents not immersed in a particular culture would not know about, or would not know how to calculate these values. Even an agent that is immersed in the culture might choose to ignore the rules associated with these values (but it would be aware of the transgression). Finally, an agent might not be able to accurately observe or compute the values (which frequently require a significant cognitive load and accurate observation of the environment). Agents might also make mistakes when planning their actions - especially in cases when they interact with agents that use a different set of values. The latter cases constitute cases of bounded rationality. We describe the design of CSSMs for an inter-cultural scenario called "Give Way" in Section 3.5.

In Chapter 4, we model the Spanish Step flower selling scam. In this scenario, a crooked seller tries to sell overpriced flowers to a prospective customer by exploiting his self-evaluated politeness and dignity, as well as his perceived peer and public image. The clients are aware that they are being cheated: the fact that the scam sometimes succeeds requires a very precise manipulation of the social sentiments by the crook. In the real world, both the seller and the client learn from this scenario, but while the seller has many opportunities to learn a social behavior strategy that leads to an occasional success, a typical victim of

the scam had usually participated in the scenario the first time. In addition, most of the victims are foreign tourists who are unsure about their estimates for the social metrics in the new environment. This scenario is modeled in Section 4.1 using the framework described in Chapter 3. By tracing two real-world outcomes of the scenario, Section 4.2 discusses how the system can explain and predict the outcomes of such social encounters.

In Chapter 5, we show how the crooked seller of the Spanish Steps scenario can learn a profitable strategy through reinforcement learning. Although the search space defined by the social calculus is large, we found that function approximation based Q-learning allows us to successfully learn efficient strategies in a relatively small number of runs. The learned strategy allows the seller to manipulate an unprepared tourist's social values of politeness and dignity, as well as his perception of the opinion of peers and the crowd.

Another way in which we extend our modeling of social interactions is to consider series of social interaction scenarios. For instance, in the Spanish Steps scenario, the public perception of the seller can evolve over longer time frames spanning multiple interactions with different clients. One of the most intriguing aspects of public perception modeling is the way in which knowledge of individual actions propagates in space and time, how interactions at different spatio-temporal locations affect each other through the public perception and how does the general public (such as a crowd of bystanders) forms and forgets a perception. The work described in Chapter 6 use an extended version of the Spanish Steps scenario that follows the interaction of a seller with multiple clients over a longer period of time. We make an effort to realistically model the public perception of the ever-changing crowd at a tourist

attraction. In Section 6.1, we discuss the mechanisms for multitasking from the point of view of the seller who tries to sell flowers to multiple clients at a time. The seller needs consider that the knowledge and beliefs propagate among the clients, influencing the outcome of the individual scenarios. We show the results of an experimental study in Section 6.2.

The human behavior can be influenced by many different factors such as conscious decisions, cognitive fallacies, psychological factors and even physiological states. Assigning numbers to social values is an inherently inexact science. However, the working assumption is that the culture enforces a more or less uniform method to calculate the sanctioned social values. This means that we can validate (and, if necessary calibrate) the CSSM model by performing a survey in which persons cognizant with the respective culture will judge the impact on the social values. In Chapter 7, we conducted a user study asking participants to rate the degree of the CSSM in a particular social scenario. This study provided us data for understanding: Identify / validate what CSSMs are relevant to a particular social interaction; Identify whether the subjects from the same culture will judge the values of CSSMs similarly; Verify whether the CSSM model is relevant in the case of scenarios involving robots that is can robot be subject of politeness or dignity CSSMs. In the chapter, we discussed only most distinctive behavior exhibited by the users for a particular CSSM. The graphical statistics of complete survey is added in the Appendix. B for future reference of data.

1.0.1   Outline and Publications

In a outline, this dissertation is organized as follows.

Chapter 1 presents the problem definitions. Chapter 2 conducts a literature review covering the background knowledge for the remaining chapters.

Chapter 3 presents a model which allows agents observing or interacting with humans to perform a calculus of social behavior by capturing cultural conventions and socially acceptable behavior models. We discuss the formal components of the model (culture sanctioned social metrics, concrete-beliefs and action impact functions).

Chapter 4 presents a detailed case study of a crooked seller who relies on the manipulation of public perception, we demonstrate the explanatory and predictive power of the model. For instance, we successfully explained behaviors which, from a utility maximization perspective would appear irrational.

Chapter 5 presents an approach for learning a profitable strategy by the seller for a scenario using function approximation based Reinforcement learning. We described the convenient mathematical framework for the action impact function (AIF) in modeling real world scenarios.

Chapter 6 presents the extension of the CSSM framework, the mechanisms for multitasking from the point of view of the seller: how can the seller interleave the actions of multiple selling scenarios? How does the knowledge and beliefs propagate among the clients of the same seller?

Chapter 7 presents the user study in which a person cognizant with the respective culture judges the impact of CSSMs in a particular social scenario.

Chapter 8 concludes the dissertation while providing few possible extensions for future work.

Results of the research work leading to this PhD thesis have been published in various conferences and book chapters. Presentations have been given at national and international conferences.

- Social calculus - Operational framework for public perception and social interaction.
  T.S. Bhatia, S.A. Khan, and L. Bölöni. Towards an operational model for the propagation of public perception in multi-agent simulation. In Proc. of 13th Int'l Workshop on Multi-Agent Based Simulation (MABS-2012), June 2012.

- Social calculus - Inter-cultural social interactions.
  T.S. Bhatia, S.A. Khan, and L. Bölöni. A modeling framework for inter-cultural social interactions. In Proc. of 2nd Int. Workshop on Human-Agent Interaction Design and Models (HAIDM-13) at AAMAS-2013, pp. 16-31, 2013.

- Social calculus - Multiple scene interactions.
  T.S. Bhatia, S.A. Khan, and L. Bölöni. Modeling the propagation of public perception across repeated social interactions. In Multi-Agent-Based Simulation XIII, LNCS 7838, 2013.

- Social calculus - Action-impact function learning.

  T.S. Bhatia, S.A. Khan, and L. Bölöni. The education of a crook: reinforcement learning in social-cultural settings. In Proc. of 13th Int'l Conf. on Autonomous Agents and Multiagent Systems, (AAMAS), 2014.

- Social calculus - Market checkpoint scenario.

  S.A. Khan, T.S. Bhatia, S. Parker, and L. Bölöni. Modeling the interaction between mixed teams of humans and robots and local population for a market patrol task. In Proc. of 25th Int'l Conf. of Florida Artificial Intelligence Research Society (FLAIRS-25), pp. 50-55, 2012.

- Validating the social calculus model using a user study.

  S.A. Khan, T.S. Bhatia, and L. Bölöni. Soldiers, robots and local population - modeling cross-cultural values in a peacekeeping scenario. In Proc. of 21st Conf. on Behavior Representation in Modeling & Simulation (BRIMS), March 2012.

- Learning action impact function for the market checkpoint scenario using genetic algorithms.

  S.A. Khan, J.A. Streater, T.S. Bhatia, S. Fiore, and Bölöni. Learning social calculus with genetic programming. In Proc. of the 26th Int'l Conf. of Florida Artificial Intelligence Research Society, (FLAIRS-26), May 2013.

CHAPTER 2
RELATED WORK

This paper is concerned with a calculus of human social-cultural behavior with the objective to provide explanatory and predictive power. We will discuss related work in three different fields:

- Models of social and cultural behavior in the social sciences - such as psychology, sociology and anthropology. Although these fields favor the form of a narrative rather than formal description, many researchers have expressed their insights in a numerical form, which can be relatively easily translated into computational models.

- Models of social behavior in formal sciences, such as mathematics and theoretical computer science. Many of these efforts formalize models originally proposed in social sciences, but there are cases where insights originally made in mathematics had been found to be applicable in social sciences. One such example is the remarkable success of network science in modeling human social behavior.

- Models of human social behavior in engineering, built with the goal of a specific application. Engineering solutions are often based on formal models or inspired by theories developed in social sciences. Nevertheless, the practical requirements of an engineering problem, such as the scarcity of available data and performance considerations some-

times led researchers to start from a blank state, and build problem-specific models based on purely engineering considerations.

It would appear that the ideal arc of ideas would be such that social scientists develop an understanding of human social-cultural behavior, which is then formalized by mathematicians. In the next step, the formal models would be used by robot enthusiast and software developers to create robots or software agents which can act appropriately in human society. In practice, however, the flow of ideas and models between these fields of human endeavour often follows a more complex trajectory. The interaction between these fields often happens at the level of awareness of problems, types of possible approaches and case studies, rather than transitioning fully formed theories or models.

2.1   Social and cultural models in the social sciences

In a way, the entirety of social sciences deals with the issue of social behavior. The goal to develop accurate models of social behavior is one of the historical traditions of sociology, tracing its origins to the work of Auguste Comte. The most prominent representative of this school of though was the structural functionalism of Talcott Parsons [8]. These efforts have been sometimes criticised as being unrealistically ambitious [9]. There are, however, a number of influential models that due to their concrete nature and focus on the details of the behavior had been especially suited towards the transition towards formalization and practical implementation.

One of the most influential models from our perspective is politeness theory, initiated by Brown and Levinson [10], and extended by many other researchers. The overall assumption is that politeness centers around the maintenance of "face" defined as the public self image of the adult human. More specifically, they define the "positive face" which refers to one's self esteem and the "negative face" which refers to one's freedom to act.

The Brown and Levinson model is often interpreted in terms of the work of Paul Grice [11] who formulated the *cooperative principle* in conversations. According to the four maxims formulated by Grice speakers in a collaborative conversation should be truthful, provide an appropriate amount of information (not too much, not too little), be relevant and avoid obscurity of expression.

Almost always, the desire to be polite (in the Brown and Levinson definition) and the desire to be cooperative (in the sense of Grice's maxims) are countervailing forces. For instance, the indirect strategy is highly polite, but leads to inefficient communication.

The Brown-Levinson model, by positing two metrics which humans want to maximize, was one of the direct influences for our approach of defining CSSMs. The most significant difference is that CSSMs are easy to collect: the intra-cultural uniformity conjecture implies that we can ask any member of the culture to evaluate them. In contrast, the terms positive and negative face do not mean anything to an untrained participant; their values must be evaluated by people with significant training. Furthermore, both the Brown-Levinson and Grice models attempt to discover the culture-independent universals in human communication. The Brown-Levinson definition of politeness does not necessarily match the definition

of politeness and indeed the desirable behavior in specific cultures. There are cultures, for instance, where direct speech is considered polite and desirable. The interpretation of the Brown-Levinson model in the context of specific cultures is a significant ongoing research topic [12, 13].

Another influential model, which specifically attempts to account for and quantitatively measure cultural differences, is the cultural dimensions theory of Geert Hofstede [14]. In the most recent publications, six dimensions are considered: (1) power distance, the acceptance of unequal distribution of power, (2) individualism versus collectivism, (3) uncertainty avoidance (4) masculinity versus femininity, a metric measuring the balance between assertiveness and competitiveness versus a focus on cooperation, human relations and quality of life, (5) long term versus short term orientation and (6) indulgence versus self-restraint. From the point of view of our model, CSSMs can be associated with one or more of these dimensions - for instance dignity has relevance to (1) and (4), while wealth to (5) and (6). Furthermore, Hofstede's analysis shows us that even if two cultures define the same set of CSSMs, they might weight these CSSMs differently in practical behavior.

2.2   Social and cultural models in the formal sciences

Formal sciences, such as mathematics, and the formal-theoretic branches of computer science and linguistics had also yielded research results which are relevant to the study of social and

13

cultural behavior. These models are distinguished by the use of mathematical formalism in their description.

The emergence of *network science* initiated by Watts' work on small-world networks [15] and Barabási and Albert's work on scale free networks [16] led to thousands of subsequent studies. The literature being very large, we can only consider several representative examples. Kottonau and Pahl-Wostl [17] studied the evolution of political attitudes in response to political campaigns - while in earlier work they studied the problem of new product diffusion. Motani et al. [18] implemented a virtual wireless social network based on the information spread in real social network such as marketplace. Gruhl et al. [19] and Adar et al. [20] analyzed the person-to-person information flow over blog-space topic sharing. Recent analysis of Twitter followers by Cha et al. [21] had shown that the influence of a user on the topic can be gained by a concerted effort over a long period of time and a large number of followers are not an assurance to fame. A significant amount of research had been directed towards the epidemic propagation of information in social networks [22, 23, 24]. While the scenarios we studied up to this moment are relatively small, our future work involves the propagation of CSSMs and CBs between scenarios, thus the progress made in this field can be an important inspiration for us.

Another formal science which provided models for social-cultural behavior is that of computational linguistics. For instance, Bramsen et al. [25] develop models which extract social power relationships between individual speakers from the language used by the speakers, relying on words and features which can be identified by natural language processing

software. Danescu-Niculescu-Mizil et al. [26] used computational linguistics to determine the type of politeness strategies (in the Brown-Levinson sense) used by Wikipedia authors in their communication logs. Again, this work provides foundation for our future efforts in automatic evaluation of CSSMs and CBs in observed scenarios.

2.3 Social and cultural models in engineering

Social models developed in engineering aim to develop artifacts such as software agents, avatars, websites or robots which take into account the social and cultural environment in which they are used. Requirements of practical applicability dominate in these fields. Engineering artifacts face additional challenges in their deployment, for instance the problem of sensing the social signals made by humans (see Vinciarelli et al. [7]).

Some of the engineering research is directed explicitly towards the practical deployment of models proposed in the social sciences.

Miller et al. [27] describes a software product called the Etiquette Engine which uses the Brown-Levinson politeness model [10] to assess the politeness in interactions involving military personnel of common culture but different rank (such as the interaction between a corporal and a major). In a follow-up work [28] the authors create a more complex model which investigates how culture (as examplified by Hofstede's cultural factors [14]) as well as politeness levels affect the way in which people react to instructions, commands or requests ("directive compliance").

Bosse et al. [29] formalizes Damasio's theory of consciousness [30], where consciousness is built up from the distinct elements of emotion, feeling and core consciousness, the latter being defined as the "feeling of a feeling".

The ubiquity of user interfaces featuring synthetic characters naturally led to the requirement that they exhibit appropriate social and cultural behaviors, such as empathy. Paiva et al. [31, 32] describe the functionality of a virtual environment called "Fear not!" which allows 8-12 year old children to witness bullying situations from a third person perspective. To model the emotions of the characters in the simulation, the system uses the cognitive theory of emotions of Ortony et al. [33]. An extension of this work is described by Rodrigues et al. [34] where the empathy model relies on the neuropsychological theories of Perceptual Action Model (PAM) [35] and the work of Vignemont and Singer on the emphatic brain [36].

Another relevant point of view is that of *social intelligence*, defined as the ability to act for social benefits. For instance, Hogg and Jennings [37] describes a model for socially intelligent reasoning for autonomous agents. The authors rely on Harsanyi's social welfare function [38] to balance the benefits to others in the course of taking an action and weight it against its own benefits.

A different view on socially responsible actions is taken by Kalenka and Jennings [39]. As before, the agents consider a joint benefit function which is a combination of the individual and social benefits. However, the authors notice that in practical scenarios, there are many actions where the societal benefit depends on another agent (or agents) taking a set of specific

actions. Without a guarantee for this collaboration, individual agents will never choose these actions. For such systems to operate in a socially responsible manner, we need a framework of *social commitments* such as the one described by Castelfranchi [40]. The authors consider a simulation of unloading trucks in a warehouse, where forklifts controlled by agents can decide to cooperate by assisting others with the unloading tasks.

The ideas behind social intelligence have found applications in a number of specific applications. For instance Grimaldo et al. [41] apply social utility functions to model the interactions between different people in a university bar, with the actors being divided between waiters and customers, the latter grouped into teachers, undergraduates and graduates.

As mobile robots are increasingly deployed in situations with human interaction partners and bystanders [42], the field of human-robot interaction [43] must increasingly consider issues of social intelligence [44].

Another relevant research direction concerns embodied agents and robots design to interact with human users in a sociable way.

The Rea system (Cassell et al. [45]) implemented an avatar which engaged in a multimodal conversation with a human user, using and understanding speech, gestures, body posture, gaze and other social signals. All these modalities contributed to the dialog between the robot and the human user, and could provide either propositional information (the content of the conversation) or interactional information (regulating the flow of the conversation).

The Kismet robot (Breazeal [46]) was developed in the context of the Sociable Machines Project at MIT. The principles behind the robot integrated theories of infant social development, psychology, ethology and evolution. The robot was able to infer emotions in the human users, and to emulate and display emotional states such as anger, fear, disgust or sorrow.

While we can learn many practical lessons from sociable robot projects, there are also several important differences. Inevitably, these projects put the robot front and center, and the social interactions are always modeled between the robot itself and the interaction partner. In contrast, our work models general purpose social interactions, with or without the active participation of robots. Another major difference is that projects such as Kismet consider theories of emotions which are actually felt by the human user (but are only emulated by the robot). In contrast, CSSMs are imposed from outside, by the culture. While Kismet models the ways humans are guided by their *emotions*, we are modeling behaviors guided by social and cultural *conventions*.

CHAPTER 3
SOCIAL CALCULUS

The fields of sociology and psychology have a rich literature of describing human behavior in specific cultural contexts. Social calculus, however, requires explicit formulas or algorithms which take as input the observable facts of a situation and specific actions, and provide an output in the form of quantitative metrics. The models developed in humanities are rarely expressed in such quantitative form. In recent years, there is an ongoing effort to *operationalize* models from sociology and psychology [27, 28, 29]. Alternatively, we can design new models of reasoning in a social-cultural context, which are informed by the sociological models, but designed from ground up to provide an implementable algorithmic framework.

This chapter introduces a formal model for the representation of human interaction scenarios, with special focus on the separation of the progress state and the full state. We also argue that to achieve predictive and explanatory power in the majority of the social scenarios, it is sufficient to consider two types of components of the state: culture-sanctioned social metrics (CSSMs) and concrete beliefs (CBs). In Section 3.2 discusses CSSMs and the way in which they depend on the culture, as well as on the subject, perspective and estimator actor. In Section 3.3 discusses CBs which model the beliefs held by an actor or the general public about salient features of the scenario, and show their dependence of the scenario, as

well as the perspective and estimator actors. Section 3.4 describe action-impact functions (AIFs), the ways in which actions taken by the actors impact the CSSMs and CBs. AIFs can in principle take arbitrary forms, but we argue that certain closed form expressions can represent both a convenient mathematical framework as well as offer sufficient fidelity in modeling real world scenarios.

3.1 Scenario modeling

3.1.1 The scenario model

The modeling of arbitrary, free format interactions between humans is clearly out of reach for theoretical models. Instead, we model specific *scenarios of human interaction* which center around the resolution of a small number of issues, have a limited number of participants and a limited time span.

Definition 1 *We call a* scenario $\mathfrak{S}$ *a tuple* $\{\mathbf{A}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \mathbf{S}, \mathcal{F}, \mathcal{P}\}$, *where:*

$\mathbf{A} = \{A_1, A_2 \ldots\}$ *is a set of* actors, *who are usually humans, although they can also be autonomous robots or software agents. In certain situations we can also introduce a* nature *actor, whose actions model the stochastic outcome of certain actions.*

$\boldsymbol{\alpha} = \{\alpha_1, \alpha_2 \ldots\}$ *is a set of distinct* action types. *A concrete action a is characterized by* $a(\alpha, A, x_1 \ldots x_n)$, *that is, by the action type, the performing actor and a list of*

parameters of arbitrary length. We denote with $\mathbf{a} = \{a_1, a_2, \ldots\}$ the (not necessarily

discrete) space of all possible actions.

$\boldsymbol{\tau} \subset \boldsymbol{\alpha}$ *is the collection of* terminal action types. *A terminal action, for any actor and*

parametrization, terminates the scenario (moves it to a terminal state).

$\mathbf{S} = \{S_1, S_2 \ldots\}$ *is the (not necessarily discrete) collection of* full states of the scenario.

$\mathcal{F}$ *is the* action impact function $\mathcal{F} : \mathbf{A} \times \mathbf{S} \times \mathbf{a} \to \mathbf{S}$. *We interpret* $S' = \mathcal{F}(A, S, a)$ *as*

the new full state of the system if actor A performs action $a(\alpha, A, x_1 \ldots x_n)$ in state S.

$\mathcal{P} : \mathbf{A} \times \mathbf{S} \to \boldsymbol{\alpha}^*$ *is the* progress function. *We interpret* $\mathcal{P}(A, S) = \{\alpha_{p1}, \ldots, \alpha_{pn}\}$ *as the*

set of action types available to actor A in state S. If the actor can perform a certain

action type, it is free to use an arbitrary parametrization of it. If in a given state no

actor can perform any action type, we call it a terminal state.

While the actions are assumed to always succeed, actions with stochastic outcomes
can be modeled through the usual game theory technique of a nature actor taking an action
after the human actor, with the nature actor stochastically either accomplishing or not the
intent of the human actor's action.

3.1.2 The progress model

The progress function $\mathcal{P} : \mathbf{A} \times \mathbf{S} \to \boldsymbol{\alpha}^*$ had been defined on the full state space of the
scenario $\mathbf{S}$. This space state is not necessarily discrete and even when it is, its size increases

exponentially with the number of variables describing the state, with the base of the exponent being the number of possible values for each variable. For instance, for the Spanish Steps scenario discussed in Chapter 4, the number of possible states has a magnitude of 10^{20} if the variables are quantized into 10 groups.

In the following, we introduce a structure which helps us analyze scenarios by observing that many human interaction scenarios are *progress-segmented*, that is, the full states can be grouped into equivalence classes with regards to the output of the progress function.

Definition 2 *We define* $\mathbf{P} = \{P_1, P_2 \ldots P_n\}$ *the collection of a finite number of* progress states. *A progress state* P *is a (not necessarily discrete) collection of full states, such that* $S \in P \wedge S' \in P \Rightarrow \forall A \quad \mathcal{P}(A, S) = \mathcal{P}(A, S')$. *The* progress state discretization *function* $PSD : \mathbf{S} \rightarrow \mathbf{P}$ *maps states to progress states.*

The progress states represent a reduction of the full state space because each progress state corresponds to multiple full states. Furthermore, there is always a discrete number of progress states while the full state space can be continuous.

Definition 3 *We will call the function* $\mathcal{P}_R : \mathbf{P} \times \mathbf{A} \rightarrow \boldsymbol{\alpha}^*$ *the* reduced progress function *and define it as* $\mathcal{P}(A, S) = \mathcal{P}_R(A, PSD(S))$.

In contrast to $\mathcal{P}$, the reduced progress function $\mathcal{P}_R$ is defined on a discrete and (usually) small space. We will also consider an even more specific class of scenarios where for every progress state only one actor is allowed to take actions.

Definition 4 *A* turn taking scenario *is a progress-segmented scenario where for any progress state P the reduced progress function* $\mathcal{P}_R(A, P)$ *is non-empty for at most one specific actor* A_t. *We say that* A_t *has the turn* in progress state P.

3.1.3 A simple example: Human Bargaining

To illustrate the model, let us consider a simple example. In the Human Bargaining scenario two actors, a seller A and a buyer B are arguing over the price of a good. The action type set contains three action types: $\boldsymbol{\alpha} = \{\alpha_O, \alpha_a, \alpha_w\}$ with the following interpretation:

α_O make an offer

α_a accept the latest offer

α_w withdraw from the bargaining

The choice of the parameters is a function of the action type, the social context and the goals and limitations of the model. If we assume that the actors are software agents exchanging numerical offers, a single parameter (the value of the offer) is sufficient. If, however, the actors are humans negotiating face-to-face many other parameters can be considered: the verbal phrasing of the offer, the politeness of the addressing form used, the tone of the voice, the body language and facial expressions which accompany the offer and so on.

The scenario can be modeled using only four progress states $\mathbf{P} = \{OA, OB, TN, TP\}$:

OA turn of A to take an action

OB turn of B to take an action

TN the bargaining had been broken with no deal

TP deal accepted

Note that this is a turn taking scenario: in progress state OA only actor A can take an action, in progress state OB only B, while TN and TP are terminal states.

In this case the reduced progress function $\mathcal{P}_R$ can be visualized as a *progress graph*, a directed graph where the nodes are progress states and the edges are labeled with the pair of an actor and an action type (see Figure 3.1).
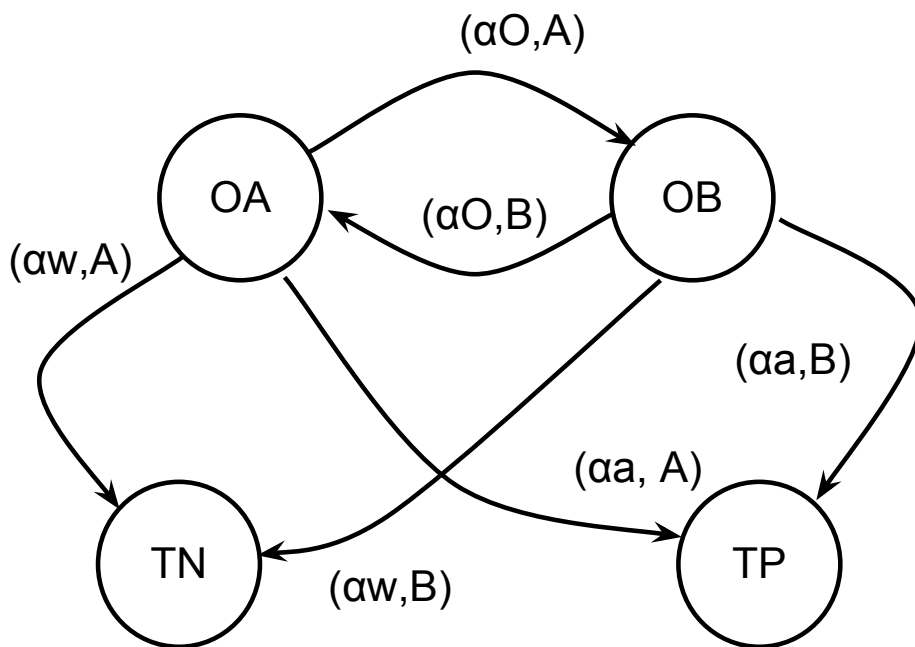
Figure 3.1: The progress graph of the Human Bargaining scenario

The progress graph is a helpful modeling tool for the knowledge engineer, but it should not be mistaken for a *full state-action graph* of the scenario. Such a graph would have full states S as nodes and fully parameterized actions $a(\alpha, A, x_1 \ldots)$ as edges. The full state-action graph is a suitable model for decision theoretic analysis - for instance, it can form the basis of a Markov Decision Process. The progress graph is not sufficient for this.

Normally, the full state includes orders of magnitude more information than the progress state. Even if A and B are software agents, the full state would have to include the pending offers, the internal valuations of the good by the actors A and B, their negotiation strategies, and possibly other factors such as their models of each other. If A and B are humans, the full state is even more complex: it might include factors such as the level of annoyance of the actors, judgment of personal dignity, the feelings of friendship or animosity against the negotiating partner, and so on. The progress state discretization function PSD stripes off all these features of the states, except the four states of the negotiation (turn A, turn B, success and failure). For instance, all the states where A has the turn will be mapped to progress state OA, regardless of the current offer, the mental states of the actors and so on.

We conclude that the full state-action graph is too large for human visual analysis, even for the most simple scenarios such as Human Bargaining. In contrast, the progress graph remains small enough for human intuition for turn taking scenarios, and even for some scenarios which do not verify the turn taking criteria at every progress state.

3.1.4 Scenario instances and social agents

Let us now discuss the relationship between a scenario, as discussed before and a concrete instantiation of it. In the Human Bargaining scenario, for instance, we talked about a buyer and seller actor. However, whenever such a scenario takes place, the buyer and the seller will be played by specific humans, let us say Jack and Jill.

We will say that the scenario instance is the combination of the scenario $\mathfrak{S}$ and a mapping of actors to social agents $SI = \{\mathfrak{S}, \langle A_1, SA_1 \rangle, \langle A_2, SA_2 \rangle, \ldots\}$ where SA_i is a social agent. Social agents maintain their own *private state* S_{SA}. The state of a scenario instance is a superset of the union of the private states of the social agents playing the actors:

$$S_{SA_1} \cap S_{SA_2} \cap \ldots \subset S_{SI} \tag{3.1}$$

This has several practical implications in the modeling. First, the scenario instance does not start with an empty state: Jack and Jill bring into the scenario their private states, which includes their personal experience, opinions, prejudices, current mood and state of mind.

If the social agents participate in a series of consecutive scenarios, the private state reached in a certain scenario instance will be carried over to the next scenario. For instance, the level of annoyance of the seller might be a significant factor in the outcome of human bargaining scenarios. If Jack had played the role of the seller in a number of scenario instances (with the buyer being played by Jane, Mary etc.) and became increasingly upset,

he will start the scenario instance with Jill in a more annoyed state, a fact which might affect his actions and the outcome of the scenario.

A different situation appears when a social actor *simultaneously* participates in multiple scenario instances (possibly, of different kind). As the private state of the social actor might be changed by action impact functions in any of these scenario instances, the implication is that the state of the scenario might be changed by actions taken place in other scenarios.

Finally, let us now discuss the nature of the social agents participating in the scenarios. The social agents might be *individuals:* humans, robots or software agents. However, in some scenarios we might consider social agents played by groups of individuals such as crowds, peer groups or families. Naturally, the private state and action impact functions of group-type social agents require special considerations.We will further increase the representational power of the model in two ways:

- We will allow *group actors* to model coordinated groups or unorganized crowds of humans. Naturally, the state and the actions taken by the group actors must be compatible with their internal organization.

- We will allow more than one scenario to be executed simultaneously, with some participants (for instance, the crowd) participating in more than one scenario. As the states of the shared participant are part of the full state of both scenarios, this will allow state information to *leak* from one scenario to another.

3.1.5 The next steps: refining the state and action impact

The progress model, as defined above, allows us to describe the general flow of the scenarios without having to consider the full richness of the state characterizing human interactions. The progress states and the reduced progress function describe the various options actors have at specific points in the scenario - but they do not have explanatory and predictive power. If we stop our modeling at this moment, we can explain what choices of actions the actors have at each point in the scenario, but we cannot neither predict what they will do, nor (in retrospect) explain why they did it.

To add such power to our model, in the following sections we take a closer look at the full state space $\mathbf{S}$ and the action impact functions $\mathcal{F}$. In particular we will make the claim that a large majority of human interaction scenarios can be modeled with explanatory and predictive power while restricting the state to only two types of values: *culture-sanctioned social metrics* attached to social agents participating in a scenario, and a small set of *beliefs about concrete facts*.

3.2 Culture-sanctioned social metrics

3.2.1 Definitions

There are many aspects of the private state of a social agent which can affect the outcome of a scenario in which it participates - this includes physical, social, cultural, psychological and

even physiological aspects. There are many examples where relatively obscure causes had significant outcomes on human interaction scenarios (see for instance the work of Kahneman and Tversky about judgement under uncertainty [47]).

To populate the private state of social agents, we are considering a set of explicit metrics, which are well specified for a given culture and can be readily estimated by the social agents. We will call these *culture-sanctioned social metrics* (CSSMs). We say that a culture *sanctions* a metric if it:

- provides a *name* for it

- provides an (informal) *algorithm* for its evaluation

- expects its members to *continuously evaluate* the metric for themselves and salient persons in their environment

- provides *rules of conduct* which depend on the metric

The CSSMs can be either tangible or intangible. Tangible metrics such as financial worth or time spent doing something can be measured by physical means (although many times they are only estimated). Intangible metrics, such as politeness, dignity, "face" or "manliness" are socially constructed, not directly measurable and depend on the specific culture. The separation between tangible and intangible metrics is often fuzzy, because even the tangible metrics such as time are interpreted by the human agents.

CSSMs are always defined by a specific culture, and the name of the metric is given in the language of the culture. Knowing the name of a metric is insufficient: it order to

be educated in a culture an individual must know the evaluation algorithm and the rules of conduct associated with it. It is not guaranteed that a given individual will *follow* the rules of conduct - however, he or she will be *aware* of the rules and their transgression.

The same name might define different metrics in different cultures. For instance, the word "dignity" has different evaluations and rules of conduct in different English speaking cultures. The dictionary translation of the word in other languages, such as "azmat" in Urdu, "pratistha" in Hindi or "méltóság" in Hungarian, can denote even more divergent CSSMs.

This being said, there are many CSSMs which appear in several cultures in identical or near identical form. There are groups of cultures with closely related metrics - for instance the cultures aligned with the Western European model, the culture of China and nations influenced by Chinese culture and the cultures of the Near East and North Africa. In addition, certain CSSMs are cross-cutting geographical, language and religious boundaries, such as the striking similarities between "cultures of honour" in places as far away as the Scottish highlands, the Bedouins of the Sahara or the Southern USA [48].

It is beyond the scope of this work to establish specific measures of similarity between different cultures. A number of established metrics in sociology can be used as a starting point for such classification (for instance, Hofstade's cultural dimensions [14]).

3.2.2 The problem of perspective

Many rules of conduct associated with CSSMs consider not only the actor's own perspective, but also the perspective of other actors in the scenario. For instance, gestures of politeness and respect are often required to be enacted such that they are visible to and noted by not only the direct interaction partner, but also by third parties. Taking this into consideration, we propose a model where a specific CSSM is identified by five parameters: `CSSM(C,M,SA,PA,EA)`, where:

- `C` is the culture which defines the CSSM and specifies its rules.

- `M` is the name of the metric, which is unique in the given culture (but different cultures might mean different metrics under the same name).

- `SA` is the *subject agent* characterized by the metric.

- `PA` is the *perspective agent*, from whose perspective the metric is evaluated.

- `EA` is the *estimator agent*, who estimates the CSSM.

The intuition about the different agents is as follows: in the estimation of `EA`, the agent `PA` believes that the value of the metric `M` for agent `SA` is equal to `CSSM(C,M,SA,PA,EA)`. There is no requirement for `SA`, `PA` and `EA` to be all different. For a CSSM to play a role in a scenario, we need that the `EA` to be cognizant of culture `C`. In addition, it is necessary for `EA` to believe that `PA` is cognizant of culture `C` (although this belief might be incorrect). It is

31

not necessary for `SA` to be cognizant of the culture (although whether he is or not might be a factor in the behavior of other actors). A specific CSSM is always part of the *private state of the estimator agent* $S_{\texttt{EA}}$.

When referring to the CSSMs in the context of a scenario instance, it is sometimes convenient to talk about the CSSM in terms of the actors (instead of the specific agents playing them). For instance, we can talk about the dignity of the seller and the politeness of the buyer. However, the CSSM is attached to the underlying social agent: it is not the dignity of a general buyer which is offended, but of Jill, who happens to be the buyer in this scenario instance. This fact is important in explaining and predicting behavior in the case of multiple sequential or simultaneous scenario instances.

In the following we will present several examples of CSSMs which illustrate that all the five parameters of the CSSM model are necessary for building a model with explanatory and predictive power.

1. *Self perspective:* the `CSSM(Western,dignity, John,John,John)` represents John's estimate of his own dignity, in the Western cultural model.

2. *Peer perspective:* the `CSSM(Western, politeness,John,Mary,John)` represents John's estimate about how Mary sees his politeness. If John cares about Mary's opinion, he will adjust his behavior in such a way that Mary's perspective will improve. Note that this value might not be identical to `CSSM(Western,politeness, John,Mary,Mary)`, that is, Mary's own opinion about John's politeness.

3. *Cross-cultural perspective:* Let us consider the case of János, a Hungarian businessman in China, who publicly admits to a business partner Chen a mistake in formulating a purchase order. This will affect `CSSM(Chinese,Face,János, János,Chen)` that is, Chen's estimate of János's own estimate of loosing face. In this context, Chen might not understand why János would do such a thing. What happens here, is that Chen is evaluating a CSSM which János does not: János is not educated in Chinese culture, and the concept of "face" as a metric is not sanctioned in Hungarian culture. Thus `CSSM(Hungarian,Face,János,János,János)` is not defined, while `CSSM(Chinese,Face,János,János,János)`, while defined, it cannot be evaluated by Janos, who does not know the Chinese culture.

Nevertheless, this CSSM can impact the outcome of the scenario: for instance, Chen might act to prevent János from loosing face, even if János is unaware of this.

3.2.3   The intra-cultural uniformity conjecture

The multiplication of possible perspectives increases the complexity of the CSSM evaluation. If we need to consider different models of evaluating the CSSMs for every social agent, the framework would have no practical utility. The *intra-cultural uniformity conjecture* states that we don't need to consider different evaluation models on the individual basis: it is enough to model them once for every culture.

Conjecture 1 *Let us consider two human actors A and B, educated in the same culture* `C` *which sanctions a metric* `M`. *Let us now consider a scenario S and a series of actions* $a_1, \ldots, a_n$, *of which both A and B are aware. The* intra-cultural uniformity conjecture *asserts that for any social agents* `X` *and* `Y`, *if before the actions we have* `CSSM(C,M,X,Y,A) = CSSM(C,M,X,Y,B)`, *than after the actions we will also have* `CSSM'(C,M,X,Y,A) = CSSM'(C,M,X,Y,B)`, *where we denote with* `CSSM'` *the values modified by the action impact functions.*

This conjecture is supported by the definition of the CSSM: the two agents have the same information and they use the same algorithm for the evaluation provided by the shared culture. We need to emphasize that the conjecture does not say that different individuals in the same culture will *behave* the same way. One social agent might follow the rules of politeness while another might not - the conjecture only says that they would both be *aware* of the rules.

An example of what the conjecture says is as follows: let us consider two Japanese persons, one of them a participant in a social situation which involves interacting with a Westerner, while the other one an outside observer. Let us now consider that the Westerner unknowingly commits an action considered impolite in the Japanese culture. The intra-cultural uniformity conjecture states that the two Japanese participants will evaluate the level of impoliteness similarly. This fact will not be changed by the fact that the Japanese might also be familiar with the Western culture.

The question might be raised: what happens if we perform a survey with respect to a supposed metric `M` and culture `C` and find that the intra-cultural uniformity conjecture

does not hold? In this case, we will need to revise our assumptions: either (a) there are actually multiple different cultures in the surveyed population or (b) the given metric is not "culture-sanctioned".

In conclusion, between two participants educated in the same culture, any difference in the evaluation of the CSSM is reduced to the perspective actor's knowledge of specific events. (Naturally, if the estimator actor itself is unaware of an event, it will automatically mean that he or she cannot assign it to the perspective actor either).

3.2.4 The problem of cognitive load

The evaluation of CSSMs is a significant *cognitive load*. Although the culture requires every actor to continuously evaluate all CSSMs for *every* salient person in the environment, in complex situations with many actors present, many actors will not be able to evaluate every possible action impact function.

Different CSSMs, actions and actors will be differently affected by the problem of cognitive load. The more complex a CSSM, the more likely that it will not be estimated. Self-perspective CSSMs are more likely to be evaluated than cross-cultural peer perspectives. CSSMs where the subject and perspective actors are random members of a crowd will be evaluated with a lower priority than CSSMs where the subject actor and/or perspective actor is the self or close peers.

Finally, the salience of the action also affects its evaluation priority. Striking actions, such as large gestures, loud voice, strong verbal expressions will raise the action's evaluation priority.

3.2.5   Numerical values of CSSMs

In order to provide us with a computational framework, CSSMs must be assigned numerical values. For tangible CSSMs this is an easy task, because they come with their concrete measurement techniques. Thus, worth will be measured in dollars or euros while time will be measured in seconds or minutes.

Things are significantly more complicated for intangible CSSMs. How do we measure politeness or dignity on a numerical scale? While the measurements of such metrics might not necessarily live up to the standards of scientific metrology as practiced in engineering, there is a significant body of work attaching numerical values to intangibles in social settings. The measurement and comparative study of emotional, cultural and social values are regularly done in the social sciences, often using graphical tools such as the interpersonal circumplex to model personality traits [49, 50]. In business and marketing settings it is sometimes important to put a numerical value of the level of politeness of salespeople or customer service [51].

While an in-depth investigation of the techniques for calibrating the numerical values of CSSMs is beyond the scope of this chapter, we will briefly outline a technique based on

the surveys of human experiment participants. The technique uses representative narratives and a combination of direct evaluation requests and keyword characterization.

Let us assume that we need to calibrate a specific CSSM attached to a given culture (for instance, politeness in Middle Eastern cultures). We start by choosing an arbitrary range such as $[0, 100]$ or $[-100, 100]$. Next, we generate a series of story snippets which are representative to the given metric. These stories must take place in locations and circumstances representative to the culture and familiar to the experimental subject. We present these story snippets in the form of written narratives (possibly enhanced with pictures or videos). After the subjects read or watch the narratives, they will be asked to:

- directly assign a numerical value to the specified CSSM (e.g. *"Please characterize the politeness of the shopkeeper on a scale of 0 to 100"*)

- characterize the metric using keywords (e.g. *"Please provide two words which characterize the shopkeeper's behavior"*)

Based on this input, we can use statistical techniques to create a numerical scale of the CSSM and to position specific keywords on the scale. The resulting scale can then be used every time we want to model a scenario which involves the specific culture and CSSM. Note that the calibration only needs to be done once per CSSM in a given culture, it does not need to be repeated for every scenario. For a concrete application of this approach we point the reader to 7

3.3   Concrete beliefs

3.3.1   Definitions

In addition to CSSMs, the behavior of actors also depends on their beliefs about certain aspects of the current scenario. Reasoning about human beliefs is notoriously difficult, both because of the very large set of possible objects of the beliefs, and because of the difficulty in emulating how humans acquire and update them. We found, however, that for many social scenarios we can achieve explanatory and predictive power by considering only a very restricted set of beliefs: those which pertain to simple binary questions which can be, in principle, unequivocally answered. Such *concrete questions* include: "Is A holding a flower?" or "Are A and B engaged in a commercial transaction?".

Let us now clarify what we mean that a concrete question can be, in principle, unequivocally answered. Let us consider an omniscient external observer, who sees every aspect of the scenario and have witnessed the scenario from the beginning. To determine whether A is holding a flower, the observer only needs to investigate the current state of the scenario. However, to determine whether the A and B are engaged in a commercial transaction, the observer also needs to look into the history of A and B's actions. For instance, if A had asked the price of the flower, B provided the price, A accepted it and B handed the flower over, the observer can deduct that A and B are engaged in a commercial transaction. Notice that the observer had to apply an algorithm to his observations to determine the answer to the concrete question: the answer is not directly in the sensing data.

In contrast to the omniscient external observer, the actors in the scenario need to work with incomplete knowledge and limited rationality. For instance, an actor X might not have witnessed the complete scenario - he only sees that the A is holding a flower, but does not know how she got it. We will call *concrete beliefs* (CBs) the beliefs maintained by the actors in a scenario with regards to the answers of concrete questions. We say that a scenario *defines* a CB if:

- there is an algorithm which an omniscient external observer could use to unequivocally answer the question underlying the CB.

- the scenario expects at least one actor to *continuously evaluate* the CB for himself and other salient actors in the scenario.

- the scenario provides *rules of conduct* which depend on the CB *or* the CB affects the calculation of CSSMs.

3.3.2 The problem of perspective

The definition of CBs has clear analogies to the definition of CSSMs, but several important differences exist. First, CBs do not depend on the culture: while the definition of politeness varies from culture to culture, the question whether a person holds a flower or not is decidable without cultural references. Instead of being tied to the culture, the CBs are tied to a specific scenario. Another difference is that while CSSMs represent the social values of a subject

actor, *e.g.* the politeness of John, the concrete question can refer to any aspect of the scenario, including inanimate entities ("is it raining?").

Putting these considerations together, we will identify a concrete belief with four parameters: `CB(SC,BD,PA,EA)`, where:

- `SC` is the scenario instance which specifies the question.

- `BD` a description of the belief (normally, through the associated question).

- `PA` is the *perspective actor*, from whose perspective the belief is evaluated.

- `EA` is the *estimator actor*, who performs the estimate and owns the knowledge.

A number of considerations discussed in the case of CSSMs are applicable to CBs as well. The CB is always part of the *private state of the estimator actor* S_{EA}. Although there is a requirement for some actors to evaluate specific CBs, this evaluation might be incomplete or incorrect due to the lack of information or cognitive overload of the actors. CBs might come with associated rules of conduct - however, these rules of conduct can be broken by the actors.

3.3.3 Values and interpretation of concrete beliefs

Representing and reasoning about beliefs has an extensive literature in fields ranging from philosophy to artificial intelligence. In the following discussion, we do not aim to contribute

new results to this field, only to discuss the relative utility of various formal frameworks for the representation of CBs. In principle, CBs can be represented using any of the formal belief representational models. However, in practice, we need to make a concrete representation choice, which needs to balance formal rigor, practical accuracy, convenience in modeling and computational feasibility. In the following we will review some of the interpretation issues involved and justify the modeling choice currently used by our team. In particular we will discuss the modeling choice involving subjective probabilities and the fair-betting interpretation, and contrast it with our choice of Dempster-Shafer theory of evidence.

One of the most frequently used formal framework of beliefs is that of a *subjective probability*, which follows the rules of Bayesian inference. One convenient interpretation of subjective probabilities is in the form of fair betting ratios. Let us consider a scenario S and the concrete question *Q: Is the seller honest?* Then, we can interpret a value of the concrete belief CB(Q,S,A,A)=0.7 as saying that A would bet \$7 against \$3 on the fact that the seller is honest. For the case of different estimator and perspective agents we can say that CB(Q,S,A,B)=0.7 means that B would bet \$5 against \$5 on the fact that A would bet \$7 against \$3 on the fact that the seller is honest.

While the fair betting ratio interpretation has significant advantages, it models an idealized rational agent, rather than the typical human assumed in our scenarios. For instance, a human might accept a \$5 against \$5 bet, but it would not accept a \$500,000 against \$500,000 bet, because this internal utility function is not a linear function of gain and loss. It is arguable, that most humans accept bets with higher stakes if they have a higher confidence

in their belief. Furthermore, humans have been shown to behave inconsistently in expressing preference about bets. This is usually shown in the terms of the Dutch Book Paradox, where people might accept as fair a collections of bets which interact in such a way that they add up to a guaranteed loss.

In the last decades a number of alternative models of beliefs had been proposed by researchers, among others possibility theory, info-gap decision theory, the Dempster-Shafer model of evidence and the transferable belief model. None of these formalisms can be positioned as a universally applicable model of human cognition, and they all have been subject to valid criticism. These models do not, in general, obey the rules of Bayesian inference, and for all of them specific examples can be constructed where they yield counter-intuitive results. Nevertheless, in many specific scenarios, these models can capture human belief maintenance better than the subjective probability interpretation.

Our current approach relies on the Dempster-Shafer [52, 53] theory of evidence as the belief representation model. This framework has the advantage of representing belief and confidence levels in a single computational model, where the current state of belief is represented by a *mass function* which assigns fractions of a mass of 1.0 to all non-empty combinations of beliefs. Incoming new evidence changes the distribution of the mass. Using the mass function, the belief in a statement can be calculated as a value bounded by two intervals, the *belief* (or *support*) and the *plausibility*. The difference between these two values represent the uncertainty associated with the belief. In our implementation, the full mass functions are part of the private state of the estimator agent S_{EA}, however, the CSSM

calculations only use the belief component. The techniques we use for updating the concrete beliefs will be detailed in Section 3.4.2.

For an example, let us consider the concrete question "Are A and B engaged in a commercial transaction?". Let us assume that the current CB of an agent is represented by the mass distribution $m(\texttt{true}) = 0.4$, $m(\texttt{false}) = 0.1$ and $m(\texttt{true,false}) = 0.5$. With these settings the Dempster-Shafer values will be $belief(\texttt{true}) = 0.4$ and $plausibility(\texttt{true}) = 0.9$.

A special consideration must apply to CBs where the perspective or estimator agent is a group agent (for instance a crowd). Naturally, different members of the crowd can hold different beliefs. One natural way to model this is to consider that each of the members contribute to the overall mass function with a fractional mass. For instance, for a crowd of 100 people, each of them will have a personal mass function where the masses add up to 0.01. For the group agent representing the crowd, the masses of different beliefs will be the sum of the individual masses held by the members.

3.4   Action impact functions

The action impact function (AIF) $\mathcal{F} : \mathbf{A} \times \mathbf{S} \times \mathbf{a} \to \mathbf{S}$ describes the way in which the state of a scenario instance evolves under the impact of a specific action performed by an actor. We interpret $S' = \mathcal{F}(A, S, a)$ as the new full state of the system if actor A performs action $a(\alpha, A, x_1 \ldots x_n)$ in state S.

Whether the AIF is created through knowledge engineering or machine learning, it is important to reduce the design space by choosing parametrized forms which retain enough flexibility to allow explanatory and predictive power, but reduce the number of parameters which must be set by the knowledge engineer or the machine learning algorithm.

In this section, we discuss the forms of AIFs which we found appropriate for modeling the majority of the human interaction scenarios we investigated. Our first simplifying step takes advantage of the fact that we reduced the state to a collection of CSSMs and CBs:

$$S = \{CSSM_1, \ldots, CSSM_n, CB_1, \ldots, CB_m\} \tag{3.2}$$

This allows us to split the AIF into a collection of functions, one for each component of the state:

$$\begin{cases} CSSM_i' &= F_i^{CSSM}(A, S, a) \\ CB_j' &= F_j^{CB}(A, S, a) \end{cases} \tag{3.3}$$

These two types of functions will take different forms as their outputs have different semantics and numerical interpretations (the measurement conventions described in Section 3.2.5 for CSSMs and the Dempster-Shafer model described in Section 3.3.3 for CBs).

3.4.1   Action impact functions for CSSMs

In order to reduce the design space of the F^{CSSM} functions we must decide on (a) the subset of the state relevant for a given CSSM, (b) the shapes of the AIF functions and (c) the parametrized mathematical forms which can represent these shapes in a convenient way.

We have seen that the state S is composed of the private states of the participating agents. The update of a CSSM in the form of $CSSM^i$ = CSSM(C,M,SA,PA,EA_x) will be kept and maintained by the estimator actor EA_x, and this actor only has access to the other CSSMs and CBs in its own private state. These will all have the form CSSM(_,_,_,_,EA_x) and CB(_,_,_,EA_x) respectively. Thus, F_i^{CSSM} will be a numerical function depending only on the CSSMs and CBs whose estimator agent is the same as the estimator agent of $CSSM^i$.

Let us now discuss the shape of these functions. A CSSM, as we have seen in Section 3.2 can represent either tangible values such as time or money, or intangible ones such as dignity or politeness. Tangible CSSMs usually have simple AIFs. For instance, if an action takes time t_a then the action will add this value to the time CSSM. If the action involves paying the sum of m_a dollars, this will decrease the wealth CSSM with the given value.

Things are more complicated for intangible metrics, whose change can be highly nonlinear and dependent on multiple factors. For instance, for expressions of dignity, we find that humans have a sensibility threshold: they ignore trifling offenses. Similarly, there is an upper saturation threshold: a level at which the offense is so big that further increasing it would not affect the dignity level. We conclude that the shape of the AIF can include

various positive or negative slopes, thresholds and saturation behaviors. Furthermore, the change in social metrics often depends on the beliefs: we are less offended by the angry voice of the interaction partner if we believe that he is right to be angry.

There are many kinds of mathematical expressions which can generate these types of shapes. Our goal is to balance computational and modeling convenience with the hope of capturing some of the essential nature behind the metrics.

Many metrics closely related to CSSMs are modeled in psychology with the assumption of certain consumable resources in the human psyche (see for instance the hypothesis of "ego depletion" [54]). In some cases, these consumables can be actually identified as physiological measures such as the blood glucose level [55]. The evolution of various phenomena under limited resources have been extensively modeled using the sigmoid shaped logistic curve $f(x) = 1/(1 - e^{-x})$, which leads to the conclusion that it is appropriate to model the CSSM AIFs as a combination of generalized logistic functions. To allow for a more flexible representation, we will rely on a parametrized version of the logistic curve. We will start with a version of Richard's curve [56], which is a logistic function parametrized with six intuitive parameters in the form:

$$Y(t) = A + \frac{K - A}{(1 + Qe^{-B(t-M)})^{1/v}} \tag{3.4}$$

In this formula, A is the lower asymptote, K the upper asymptote, B the growth rate, while v, Q and M are parameters which affect the location and rate of maximum growth of the function. The six parameters allow for considerable freedom in the specification of the shape of the sigmoid function, but they also provide more detail than the requirements of

our problem domain. Thus, we chose to reduce the number of parameters by only keeping as variables K for the upper asymptote, M for the location of largest growth and B for the growth rate. The other values will be fixed at $A = 0$ and $Q = v = 1$. We will call this 4-parameter function the *logistic component* of the AIF:

$$L(x, K, M, B) = \frac{K}{1 + e^{-B(x-M)}} \tag{3.5}$$

With these preliminaries, we define the *logistic canonical form* (LCF) of the AIF functions as follows:

$$F_i^{CSSM} = \sum_k \left(\prod_l L(x_{kl}, K_{kl}, M_{kl}, B_{kl}) \right) \tag{3.6}$$

where x_{kl} is either the constant 1, an arbitrary parameter of the action, a CSSM or a CB. All the CSSMs and CBs which appear in the formula must have the same estimator as CSSM^i. When some of the logistic components recur in more than one term, we will sometimes write an AIF more compactly by factoring them out.

We found that the shapes we identified as necessary for the AIFs (positive or negative slopes, thresholds and saturation behaviors) can be achieved as a sum of a small number (typically one or two) appropriately characterized logistic components. Figure 3.2 shows four examples of such function shapes achieved with at most two logistic component terms: a sigmoid shape, a step function shape, a linear slope and a multi-plateau shape with two saturation plateaus.
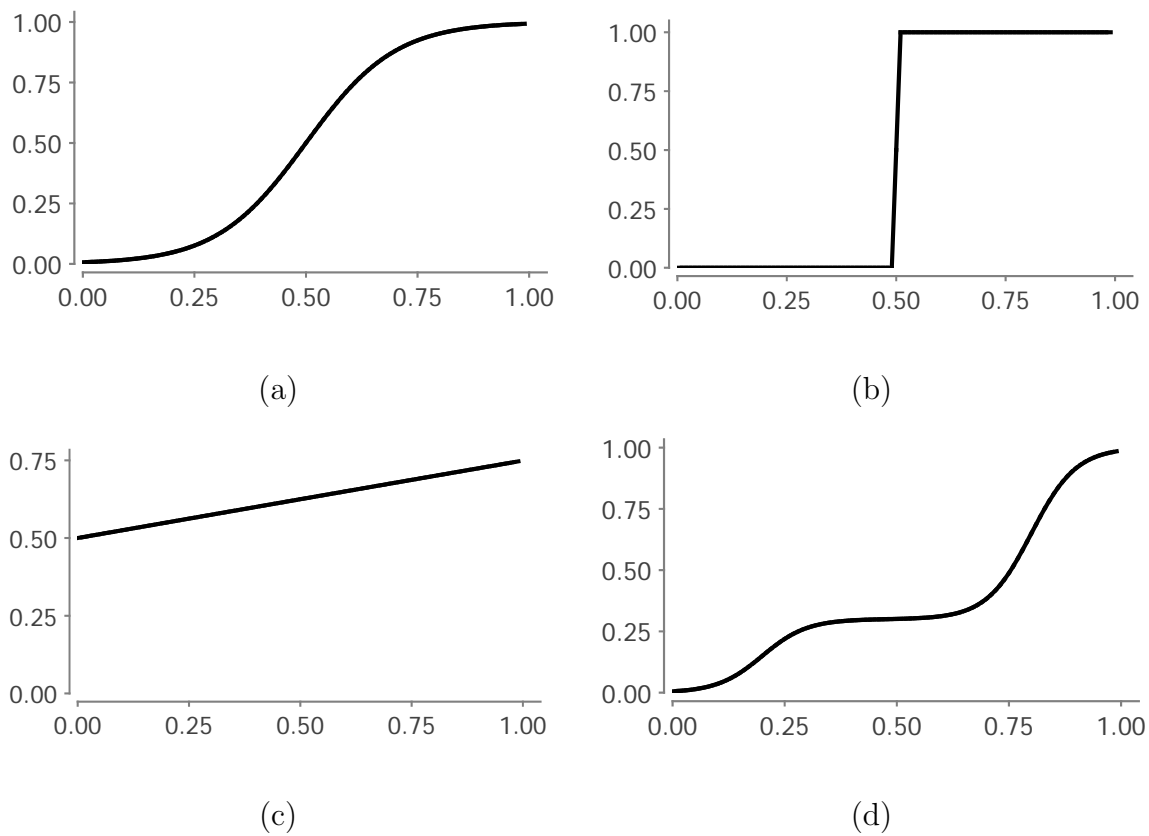
Figure 3.2: Implementing different AIF functions shapes using sums of logistic components. (a) Sigmoid shape L(x,1,0.5,10) (b) Step-function shape L(x,1,0.5,1000) (c) Linear growth shape L(x,5,0,0.2) + L(x,-2,-10,100) (d) Multi-plateau shape L(x,0.3,0.2,20) + L(x,0.7,0.8,20).

3.4.2 Action impact functions for CBs

The action impact functions for CBs encapsulate the way in which the beliefs of the agents change as a result of *evidence* provided by witnessed events and actions. As discussed in

Section 3.3.3, we will make the assumption that our representation model is the mass function of the Dempster-Shafer theory of evidence.

As the concrete questions underlying the CBs are binary, the CB can be characterized by the mass function values for $m(\texttt{true})$ and $m(\texttt{false})$. As $m(\emptyset) = 0$ by definition, we will have $m(\texttt{true } or \texttt{ true}) = 1 - m(\texttt{true}) - m(\texttt{false})$.

An evidence arriving in the form of new information received from an action performed by an agent other than the estimator or and event will also have a similarly defined mass function m_e. Actions taken by the estimator agent itself will never impact its own CBs. As the agent is free to choose its own action, the choice of the action never represents new information.

The new belief value will be given by Dempster's rule of combination (the *conjunctive merge*):

$$m'(A) = \frac{1}{1 - K} \sum_{B \cap C = A \neq \emptyset} m(B) \cdot m_e(C) \tag{3.7}$$

where $A, B, C \in \{\texttt{true}, \texttt{false}, (\texttt{true } or \texttt{ false})\}$ and

$$K = \sum_{B \cap C = \emptyset} m(B) \cdot m_e(C) \tag{3.8}$$

3.4.2.1 Modulating the evidence mass function depending the function parameters

The secret to the appropriate use of this framework is the choice of the functions which describe how much weight we put on the evidence.

Decisive evidence: Naturally, there are situations where we have evidence which decides the value of the CB decisively:

$$\{m_a(f) = 1, m_a(t) = 0, m_a(f \vee t) = 0, m_a(\emptyset) = 0\}$$

Circumstantial evidence: This means usually that the agent assumes that certain actions are more likely to happen if the value of the CB is in a certain way.

Uncertainty about the witnessed action: If the estimator agent has an uncertainty about the witnessed action, it might not consider it at full strength evidence, but it will dial back the strength of the evidence accordingly

$$\{m_{ua}(f) = u \cdot m_a(f), m_{ua}(t) = u \cdot m_a(t), m_{ua}(f \vee t) = (1 - u) \cdot m_a(f \vee t), m_{ua}(\emptyset) = 0\}$$

Partial witness of group agents: A case very similar to the uncertainty case is when the estimator agent is a group agent (such as a crowd) and only a certain percentage of the crowd witnessed the action. We will model this by setting the uncertainty value to the percentage of the crowd which actually witnessed the action.

Note that this model is a reasonable expression of the crowd behavior only as long as the crowd can be modeled as a single actor. If for instance, the crowd sharply splits along their

beliefs, the groups thus created will behave differently and this must be represented through two actors in the scenario.

3.5 An example of a inter-culture scenario

Let us now illustrate the way in which the model described in the previous sections can be used to model a simple scenario involving multi-cultural interaction called as "The Give Way scenario". This scenario involves two agents A and B approaching simultaneously a door. We assume that the agents are humans, potentially of different cultures, who can have various ages, gender and social status. The scenario also generalizes to situations where one of the agents is a robot.

For each of the agents, there are three different resolutions (1) enter the door first, (2) open and hold the door to the other agent and (3) give way to the other agent to enter first. We assume that the agents do not know each other, they might act under different cultural assumptions, that is they have different CSSMs, with different update rules and associated social requirements. A further complexity can be considered if the scenario happens in the view of the public, in which case the agents also need to consider their estimates of the beliefs of the crowd, in forms of CBs.

Table 3.1: The action types of the Give Way scenario.

Action type	Description
$\alpha 1, \alpha 4$	open the door
$\alpha 2$	enter the door
$\alpha 3$	give way to others

Let us now proceed to model the Give Way scenario using the framework developed in the previous sections. The scenario can be modeled with action types as shown in Table 3.1. The progress graph, where the nodes are progress states and the edges are labeled with the pair of an actor and an action type is shown in Figure 3.3). The scenario begins with the start state SS and continues until one of the agents perform the action $\alpha 2$ to reach the terminal state TN.
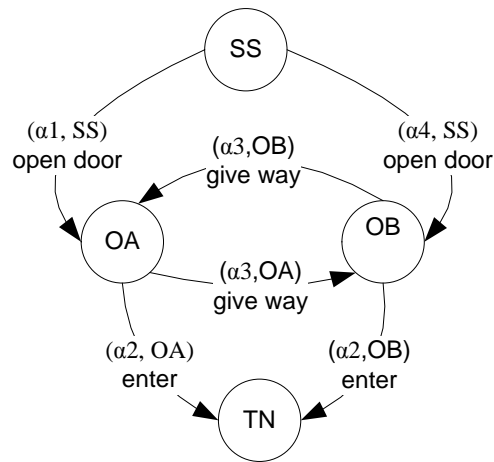
Figure 3.3: The progress graph of the Give Way scenario

Let us now consider the CSSMs which determine the behavior of the agents in the scenario. Depending on the cultural background of the agents, different set of CSSMs are evaluated. We will consider agent's of two different cultures, Western and Indian. For the purpose of this paper, we will assume that the two participants are of the same gender and they do not have a significant difference of social rank. By and large, politeness considerations in Western culture require the agents to give way to the peer (although this requirement is frequently ignored). In Indian culture, giving way is considered an ineffectual, wimpish behavior.

Thus our model will consider three CSSMs, one concrete (time), and two intangible (politeness and wimpiness). The time T is the amount of time spent on the current scene measured in seconds. Every time taking action $\alpha 3$ by agent imposes a penalty of 5 seconds. In general, agents avoid wasting time.

The politeness is the conformance to the perceived social norms of speech and gestures. Both Western and Indian cultures have the definition of politeness, but there are different definitions associated with them, which translate into different action impact functions. Giving way in the Western culture is considered polite behavior. In the Indian culture however, giving way to a stranger does not impact the perception of politeness. We will consider the private, peer and public politeness aspects: *CSSM(Western, politeness, A, A, A)*, *CSSM(Western, politeness, A, B, A)* and *CSSM(Western, politeness, A, Crowd, A)*.

Wimpiness is the degree of lack of confidence and courage in a person to take initiative. Again, both cultures have definitions for this metric. However, giving way to a person of equal

rank does not impact perception of wimpiness in the Western culture, however, it does in the Indian one. We will consider the private, peer and public wimpiness aspects *CSSM(Indian, wimpiness, A, A, A), CSSM(Indian, wimpiness, A, B, A)* and *CSSM(Indian, wimpiness, A, Crowd, A)*.

To achieve E&P power, the analysis of the scenario needs to consider two concrete beliefs, concerning the culture of the two agents, and we need to consider this from the perspective of each other and, potentially, of the crowd. Naturally, *CB(GiveWay, Is-A-an-Westerner, A, A)* is a fixed value, because normally A would know whether he is a Westerner or not. On the other hand, *CB(GiveWay, Is-A-a-Westerner, B, B)*, representing B's belief whether A is an Indian, and *CB(GiveWay, Is-A-a-Westerner, Crowd, Crowd)*, representing the crowd's belief whether A is an Indian are values whose calculation contributes to the E&P power of the model.

Let us now illustrate through several examples the way in which the model traces the evolution of the CSSMs for agents in different cultures. We have modeled the Give Way scenario using our framework, and we traced the evolution of CSSMs for four different sequences of events, each of them representing a different path through the scenario:

$$\text{SS} \xrightarrow{\alpha 1} \text{OA} \xrightarrow{\alpha 3} \text{OB} \xrightarrow{\alpha 2} \text{TN}$$

$$\text{SS} \xrightarrow{\alpha 4} \text{OB} \xrightarrow{\alpha 2} \text{TN}$$

$$\text{SS} \xrightarrow{\alpha 4} \text{OB} \xrightarrow{\alpha 2} \text{TN}$$

$$\text{SS} \xrightarrow{\alpha 1} \text{OA} \xrightarrow{\alpha 3} \text{OB} \xrightarrow{\alpha 2} \text{TN}$$

These scenarios, however, lead to different perceptions and social metrics depending of the culture of the participating humans. We will describe two experiments with different outcomes and cultural backgrounds of the participants.

Experiment 1: we consider that both agents A and B belong to the Indian culture. In this case neither of them consider the politeness as a CSSM. As both parties have the same culture, their peer perceptions are a good approximation of the opponents self perception, *i.e. CSSM(Indian, wimpiness, A, A, A)* ≈ *CSSM(Indian, wimpiness, A, A, B)*. This allows the agents to make a reasonable prediction of the opponent's actions.

Let us assume that agent B arrives at the door before A, and simply moves on without being polite and giving way to A. The *CSSM(Indian, wimpiness, B, B, B)* will be lowered, while *CSSM(Indian, wimpiness, A, A, A)* will not be affected. In colloquial terms, A can feel himself as a efficient and non-wimpy person, and this can explain its behavior, and can be used to predict it.

Experiment 2: let us now consider an experiment in which we don't know the culture of agents A and B, however, we know that the culture of the onlookers (modeled as the Crowd agent) is Indian. If we don't know what culture the agent's belong to, we can simply not trace any CSSMs and CBs for it. We can, however, trace the crowd's belief. Let us consider a scenario where A approaches the door and opens it to agent B. Let us now see how the crowd can reason about this. If A is a westerner, than his politeness level will increase *CSSM(Western, politeness, A, A, Crowd)*. On the other hand, if A is an Indian, his wimpiness will increase *CSSM(Indian, wimpiness, A, A, Crowd)*. As the same action

appears rational for a Westerner, but irrational for an Indian, the crowd will treat this occurrence as an evidence which increases *CB(GiveWay, Is-A-a-Westerner, Crowd, Crowd)*.

CHAPTER 4
SOCIAL CALCULUS - THE SPANISH STEPS SCENARIO

4.1 Scenario Modeling

4.1.1 Informal description

We consider a flower selling scam, perpetrated at many tourist sites in Italy (and probably at many other popular destinations around the world). We have seen the most aggressive example of this scam at the Spanish Steps in Rome. The intention of the seller is to pressure a client (typically a woman or a romantic couple) to purchase a flower at an inflated price. A typical interaction unfolds as follows:

- The seller offers a bouquet of flowers to the client. The client declines to purchase.

- The seller offers a single flower, relying on gestures implying that it is a gift. If the client refuses to take the flower, he repeats the offer several times, pushes the flower into the client's hands, or inserts it into her bag.

- The seller waits an amount of time at some distance from the client. During this time, the client gets used to the received gift, takes a picture with it or puts it in her bag.

- The seller approaches the client and requests payment, relying on visual signals (rubbing the pointing finger and thumb together).

- The client repeatedly attempts to return the flower while the seller refuses to take it. The action concludes by either the client paying or by escalating her verbal efforts to return the flower until the seller decides to take it back.

The Spanish Steps flower scam, despite being physically simple, is based on a series of complex decisions. It is, at its roots, a negotiated commercial transaction, which, however, is initiated by a *deceit* – the implication that the flower is a gift. The deceit is facilitated by the *blocking of the normal channels of communication* – the seller is usually a good speaker of several languages, but fakes reduced communication ability to position the deceit as a misunderstanding. The successful conclusion of the scam relies on the *manipulation of the public perception:* the client needs to have the impression that everybody around believes that she agreed to buy the flower.

Explaining and predicting the behavior of the participants is not necessarily easy even for the human observer. Why some clients accept to pay for the flower, well knowing that they are cheated? Conversely, why does the seller, occasionally, give up, without pushing the selling process to the extremes? Neither question can be answered based on the assumption of a narrowly defined wealth-maximizing rational agent.

4.1.2   Actors, progress graph and action types

Let us now model the scenario using the proposed framework. In its full setup, the scenario has three individual actors: the `Seller`, `Client`, `Spouse` and a group actor, the `Crowd`.

Out of these actors, only the `Seller` and the `Client` take actual actions. However, the `Spouse` and the `Crowd` influence the outcome of the scenario by being the perspective actors in CSSMs and CBs considered by the active actors.

The two active actors can take actions belonging to the 16 action types listed in Table 4.1. For some of these actions we also need to consider the parametrization. $\alpha 8$ and $\alpha 10$ are actions involves verbal and gestures for declining a gift and attempting to return the flower respectively. They are parameterized by their "loudness" x which determines how many onlookers will overhear the transaction and their "offensiveness" y which influences the way in which the action impacts the politeness of the actor and the dignity of the target. For action type $\alpha 13$, which involves the `Seller` waiting without taking any action, the parameter is the length of the wait t.

Table 4.1: The action types of the Spanish Steps scenario

Action type	Description
$\alpha1$	offers flowers to sell
$\alpha2$	accepts to buy the flowers
$\alpha3$	pays for flowers
$\alpha4$	declines buying flowers
$\alpha5$	offers flower as gift
$\alpha6$	forces gift
$\alpha7$	accepts flower as gift
$\alpha8$	declines gift (x,y)
$\alpha9$	throws gift
$\alpha10$	attempts return of flower (x,y)
$\alpha11$	declines return of flower
$\alpha12$	accepts return of flower
$\alpha13$	waits (t)
$\alpha14$	requests payment
$\alpha15$	gives up interaction
$\alpha16$	concedes gift
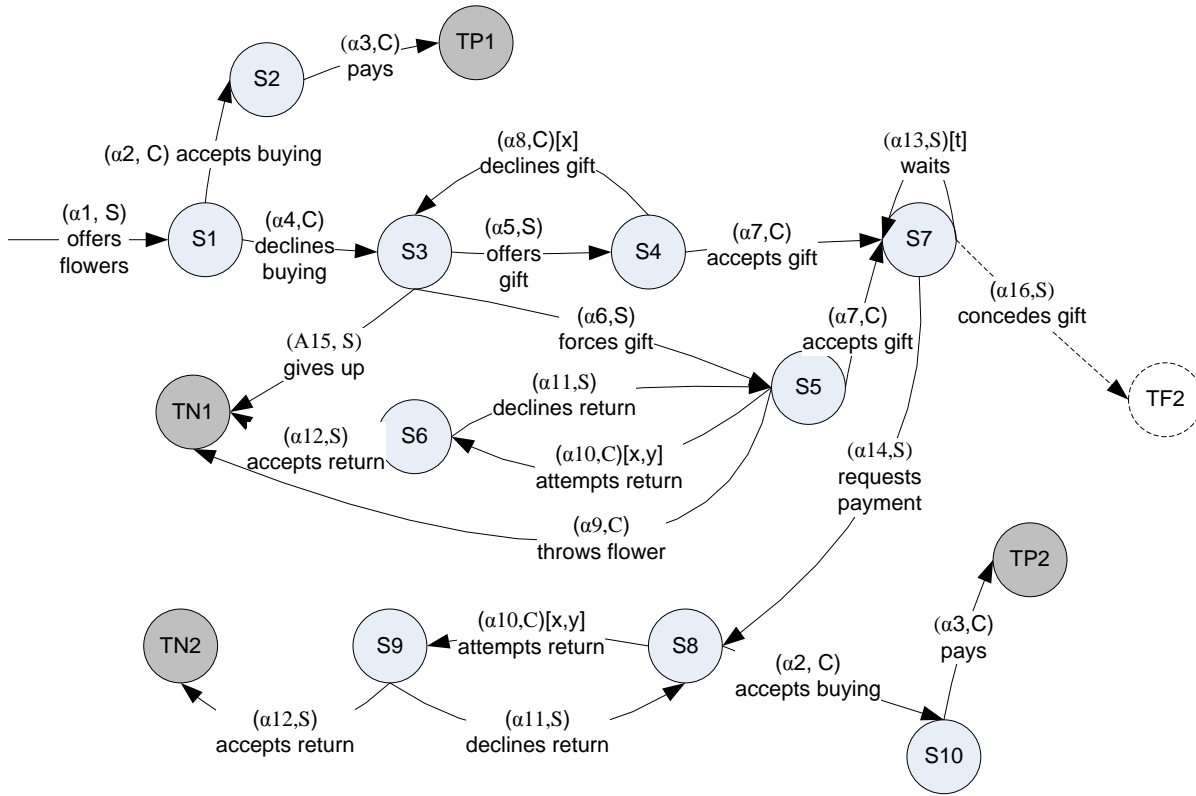
Figure 4.1: The progress graph of the Spanish Steps scenario.

The Spanish Steps scenario is a turn taking scenario, and it can be represented with the progress state graph shown in Figure 4.1 which has 10 non-terminal progress states and 4 terminal progress states.

The full state space of the scenario, depending on the degree at which we choose to model the state of the client and the seller, can be very large. The progress state discretization function PSD we use here stripes off all the internal state of the client and the seller. For instance, progress state S9 represents a situation where the client had just tried to return the flower. This progress state groups a large number of possible full states - from states in

which the client is mildly amused to states in which she feels angry, humiliated, embarrassed at various degrees, as well as possible combinations of these.

4.1.3   CSSMs in the Spanish Steps scenario

The metric of financial wealth is the central concern of every financial transaction. However, as we have seen, taken by itself, the assumption of maximizing financial wealth cannot explain or predict the behavior of the actors in the Spanish Steps scenario. In the following, we will consider a collection of CSSMs and CBs that allow us to model the scenario with explanatory and predictive power.

The first step is to decide the culture (or cultures) within which the scenario takes place. Although the real-world scenario involves people belonging to different nations – the clients and crowd are composed of usually European, American and Japanese tourists, while the sellers are usually South Asian – we found that the explanation of the seller and client actions does not require the assumption of multiple cultures with incompatible or differing metrics. Thus, we will assume that the scenario takes place in its entirety in the `Western` culture.

Naturally, as CSSMs have five parameters, the number of possible CSSMs is very large. However, many of them can be eliminated as either impossible to estimate by the specified estimator, or irrelevant to the scenario in the sense that the actions of the actors are not affected by it. Another way to reduce the modeling effort is to identify CSSMs which

have identical values throughout the scenario. The intra-cultural uniformity assumption will be very useful here, as all the metrics will be evaluated the same way by all evaluator agents, provided that they witnessed exactly the same set of actions and events.

We will use four metrics: two concrete ones (`Worth` and `Time`) and two intangibles (`Dignity` and `Politeness`). In the following we briefly describe the relevant CSSMs based on these metrics, their methods of calculation and associated rules of conduct.

`Worth:` is the sum of the financial worth of the person, measured in real-world currency. Persons in general will try to increase their financial worth. For the Spanish Steps scenario we assume that the seller and the client each consider only their own personal worth. There are scenarios where the estimation of the worth of the interaction partner is necessary for accurate modeling: for instance, in a "Giving Money to a Beggar" scenario.

```
CSSM(Western,Worth,Seller,Seller,Seller)
```

```
CSSM(Western,Worth,Client,Client,Client)
```

`Time:` is the amount of time spent in the current scenario measured in seconds. Persons, in general, will avoid wasting time. Again, we assume that the seller and the client only consider their own time spent. If the seller deals with one client at a time, these values will be the same:

```
CSSM(Western,Time,Seller,Seller,Seller) =
```

```
CSSM(Western,Time,Client,Client,Client)
```

`Dignity:` in Western culture is associated with the degree of respect the person receives from interaction partners or the degree of self-respect he shows. An insult decreases the dignity of a person. A person will feel insulted if the communication partner uses rude language, or if he perceives that he has been lied to. The dignity of a person is also affected by his own actions: for instance, an excessive emotional display decreases the dignity of the person. It is considered undignified to renege on a promise (for instance, to not fulfill an accepted commercial transaction).

As Western culture requires persons to maintain their dignity, the metric affects the decisions of the actors in the scenario. The client evaluates his own dignity from his own perspective, from the perspective of the spouse and the perspective of the crowd. These values are also estimated by the seller. As the seller sees all the relevant actions and understands Western culture, his estimate of the client's dignity will be the same as the client's own estimate.

`CSSM(Western,Dignity,Client,Client,Client) =`

    `CSSM(Western,Dignity,Client,Client,Seller)`

`CSSM(Western,Dignity,Client,Spouse,Client) =`

    `CSSM(Western,Dignity,Client,Spouse,Seller)`

`CSSM(Western,Dignity,Client,Crowd,Client) =`

    `CSSM(Western,Dignity,Client,Crowd,Seller)`

Modeling the dignity of the seller presents an interesting challenge. We might say that the seller, engaged in a deceitful selling maneuver, does not care about his own dignity or at least values it much less than financial gain. An alternative explanation would be that the seller has a different culture and thus applies a different metric. Nevertheless, even if he does not care about his own dignity in the Western definition of the metric, social pressure obliges him to consider his dignity from the perspective of the crowd. This value can also be estimated by the client:

```
CSSM(Western,Dignity,Seller,Crowd,Seller) =

    CSSM(Western,Dignity,Seller,Crowd,Client)
```

`Politeness:` in the context of the Western culture is encoded in a set of rules governing acceptable forms of speech and gesture in specific circumstances. We are considering here a relatively narrow definition of the politeness of speech forms and gestures. This is a more restricted and specific interpretation than, for instance, positive face in politeness theory [10] which tries to account for a wider range of phenomena across culture.

A person decreases his politeness metric if he uses rude language, loud voice or indecent or threatening gestures. It is considered impolite to decline a gift or to insist on an issue in the face of the refusal from the interaction partner. Western culture requires persons to maintain a positive politeness in the perception of the self, as well as from the perspective of peers and crowd. In estimating politeness Western culture also takes into account whether the interaction partner "deserves" politeness based on his recent actions. For instance, rude language addressed to a crooked seller has a smaller impact on the client's politeness metric.

Similarly to other cultures, the Western definition of a polite speech or gesture depends on the relative power distance between the interaction partners. In addition, the estimation of Western politeness also considers whether the interaction partner "deserves" politeness based on his recent actions. The rules associated with the politeness metric in different cultures vary in this respect. For instance, the cultures of Japan and Korea have more specific rules for adapting to power distance, but are less likely to grant exceptions on the basis of an undeserving interaction partner.

The politeness metric is taken into account at several action choices. The client's decision to accept the flower in state S4 is influenced by his self perception:

```
CSSM(Western,Politeness,Client,Client,Client)
```

At progress states S8 and S9 however, the client knows that he is being cheated, so his rudeness towards the obviously crooked seller will not affect his own politeness metric. However, he still needs to worry about the perception of the crowd and his spouse who might not consider the seller crooked:

```
CSSM(Western,Politeness,Client,Crowd,Client)
```

```
CSSM(Western,Politeness,Client,Spouse,Client)
```

The metric of politeness is also relevant to the seller, who must care about his own politeness as perceived by the crowd:

```
CSSM(Western,Politeness,Seller,Crowd,Seller)
```

This fact is also known by the client, who can approximate this value with:

```
CSSM(Western,Politeness,Seller,Crowd,Client)
```

4.1.4 CBs in the Spanish Steps scenario

The next step is to determine a (preferably small) set of concrete questions and associated CBs which influence the behavior of the actors in the Spanish Steps scenario. We find that we only need to consider two questions: `Q-Gift` and `Q-Agreed`.

`Q-Gift: Is the flower a gift?`

This question is unequivocally answerable by the seller (he knows it is not) so we have:

```
CB(S,Q-Gift,Seller,Seller) = 0
```

However, the value for the client `CB(S, Q-Gift,Client,Client)` has a significant impact on whether he will accept the flower or not in progress state S3. If the flower is a gift and he declines it, the client will incur a penalty in politeness. On the other hand, it is not impolite to decline a commercial transaction. Accordingly, the seller is interested to ensure that in state S3 we have a high value for `CB(S,Q-Gift,Client,Seller)`, a value which approximates the client's own belief. The value `CB(S,Q-Gift,Client,Client)` will be set to 0 at the moment when the seller asks for money, and from this point on, the CB does not influence the decisions of the actors.

Let us now move on to the second relevant question:

Q-Agreed: Has a commercial transaction been agreed upon?

The answer to this question is actually clear for both the self and reciprocal CBs of the client and the seller, as they both know that no commercial transaction took place. As a note, one could imagine a scenario where the client might be confused whether he had actually agreed to a transaction without really noticing it. However, this would not normally happen in this scenario: the client knows that he is being cheated.

```
CB(S,Q-Agreed,Seller,Seller) = 0

CB(S,Q-Agreed,Client,Seller) = 0

CB(S,Q-Agreed,Seller,Client) = 0

CB(S,Q-Agreed,Client,Client) = 0
```

These values being known and constant, they do not impact the actions of the actors. However, the perspective of the crowd estimated by the client `CB(S,Q-Agreed,Crowd,Client)` is relevant to the behavior of client. If this value is high, the action $\alpha 10$ is perceived by the crowd as reneging an agreed upon transaction, while if it is low, they judge it to be a confrontation with a crooked seller. The seller must thus act to bring `CB(S,Q-Agreed,Crowd,Client)` $\approx$ `CB(S,Q-Agreed,Crowd,Seller)` to a high value.

Notice the importance of a passive actor (the crowd) in the scenario. The Spanish Steps scam would rarely succeed on an empty street. The presence of the crowd, even without taking any active action, changes the dynamics of the scenario by serving as a perspective actor for the dignity and politeness CSSMs and the `CB(S,Q-Agreed,..,..)` values.

4.1.5 CSSM action impact functions

In the following we illustrate some of the representative CSSM action impact functions, moving from simpler to more complex AIFs. These AIFs had been knowledge engineered as follows. We started with an informal description of the impact in natural language. Then, we separated the parameters of the action and identified the ways in which they change the CSSM (step functions, linear dependency, single or multiple plateaus). Then, for each of these dependencies we chose appropriate logistic components as seen in Figure 3.2, and adjusted the parametrization until it matched the natural language description. Finally, we combined the components to obtain the appropriate multi-variable AIFs.

4.1.5.1 Worth of the client at $\alpha 3$ (client paying for the flower)

We assume that the cost of the flower is 5€ (we do not model bargaining for the price). Thus, if we denote with $v =$ CSSM(Western, Worth, Client, Client, Client) the value before the action, and with v' the same value after the action we have $v' = v - 5$.

4.1.5.2 Time of crowd at $\alpha 13$ (wait time t before asking for money)

Naturally, the time passes the same way for all the actors, independently of perspective. The CSSM of interest is $v =$ CSSM(Western, Time, Crowd, Client, Seller)= CSSM(Western,

69

`Time, Crowd, Client, Client`) because this is the value which impacts the evolution of the belief of the crowd that a commercial transaction had been agreed upon `CB(S, Q-Agreed, Crowd, Client)`.

In this case, we simply add the parameter to the time value $v' = v + t$.

4.1.5.3   Impact of $\alpha 10$ on the self-perceived dignity of the client

Let us now consider the impact of $\alpha 10$ which represents the attempt to return the flower on the self-perceived dignity of the client $v = $ `CSSM(Western, Dignity, Client, Client, Client)`. The action is parameterized by the parameters x (loudness) and y (rudeness). We calibrate the numerical values of these parameters on the scale of $[0,1]$ using the keywords in Table 4.2. We are using common sense values for the loudness. Note that, strictly speaking, the loudness can be matched to physically measurable sound pressure values, but this is less useful in developing the AIF than the intuitive metrics developed here. The low values of the rudeness parameter (0.0-0.6) are mapped to the mitigation level of speech. Higher values of rudeness involve insulting language and threats of physical violence.

Using these values, the change in dignity level can be expressed in the logistic canonical form as:

$$\Delta v = L(b, 1, 0, 4) \cdot \big(L(x, -10, 1.2, 15) + L(y, -1, 0.95, 15)\big)$$

Figure 4.2 shows the evolution of this value function of the belief and rudeness values, for the loudness parameter fixed at $x = 0.5$. Note that only high values of rudeness $(y > 0.5)$

affect the dignity of the client. Furthermore, the higher the belief that the action constitutes a reneging on an accepted transaction, the higher the impact on the dignity. Nevertheless, a certain loss of dignity occurs even if the action happens in front of a seller perceived as crooked ($b \approx 0$).
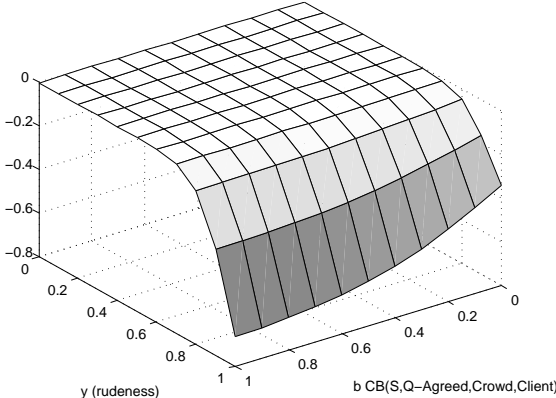

Figure 4.2: The impact of action $\alpha 10$ on CSSM(Western, Dignity, Client, Client, Client) function of b and y for a fixed value of x=0.5

4.1.5.4   Impact of $\alpha 10$ on the estimated public perception of client politeness

Let us now consider the impact of $\alpha 10$ which represents the attempt to return the flower on the self-perceived politeness of the client $v = $ CSSM(Western, Politeness, Client, Crowd, Client). The action is parameterized by the parameters x (loudness) and y (rudeness). We calibrate the numerical values of these parameters on the scale of [0,1] using the keywords in Table 4.2. We are using common sense values for the loudness. Note that, strictly speaking, the loudness can be matched to physically measurable sound pressure values, but this is less

useful in developing the AIF than the intuitive metrics used here. The low values of the rudeness parameter (0.0-0.6) are mapped to the mitigation level of speech. Higher values of rudeness involve insulting language and threats of physical violence.

Table 4.2: Intuitive keywords for calibrating the parameters of action $\alpha 10$

Value	x (loudness)	y (rudeness)
0.0	no sound	undetectable
0.1	whisper	indirect request: hint
0.2	urgent whisper	preference
0.3	subdued speech	query
0.4	speaking voice	direct request: suggestion
0.5	authoritative tone	obligation
0.6	loud voice	command
0.7	yell	generic foul words
0.8	shout	targeted offense: eg. ethnic slur
0.9	scream	
1.0	shriek	threat of physical violence

The perceived politeness can be either increased (for low values of y) or decreased (for high values of y). A louder voice can amplify the negative impact of rudeness, but it will not increase the politeness of mitigated speech. Furthermore, the impact Δv will depend on the belief of the crowd with regards to whether the action involves reneging on an accepted

transaction or whether it is the justifiable reproach addressed to a crooked seller, a value captured in the concrete belief $b =$CB(S, Q-Agreed, Crowd, Client). The higher the belief that a commercial transaction has been agreed upon, the more negative impact the rudeness of the client will have on his perceived politeness. If the public perceives the seller as crooked, the rudeness of the client will be perceived as a justifiable self-defence, and his perception will not suffer. On the other hand, the positive impact of polite behavior improves the metric regardless of the value of b (one can be polite with a crooked seller).

Denoting with $v =$ CSSM(Western, Politeness, Client, Crowd, Client), we have an AIF which can be modeled with the following logistic canonical form:

$$\Delta v = \big(L(y, -0.8, 0, 15) + L(1, 0.8, -100, 100)\big) \cdot$$
$$\big(L(x, 50, 0, 0.08) + L(1, -25, -100, 100)\big) +$$
$$L(y, -1, 0.95, 15) \cdot L(b, 1, 0.65, 8) \cdot$$
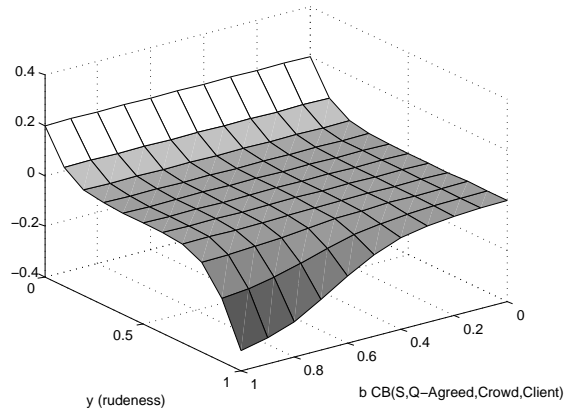$$\big(L(x, 50, 0, 0.08) + L(1, -25, 100, 100)\big)$$

Figure 4.3: The impact of action CSSM(Western, Politeness, Client, Crowd, Client) function of b and y for a fixed value of x=0.5

Figure 4.3 shows that the evolution of Δv function of the b and y values for a fixed value of $x = 0.5$ indeed matches the informal description we provided above.

4.1.6 CB action impact functions

In our model, actions affect the concrete beliefs through the application of the Dempster-Shafer conjunctive merge between a belief mass distribution representing the current belief and a belief mass distribution describing the weight of the evidence. To correctly track the evolution of the CBs we must associate a (possibly parameterized) belief mass distribution to every action.

74

Table 4.3: Mass functions of evidence for `CB(S, Q-Gift, Client, Client)`

Action	{T}	{T,F}	{F}
$\alpha 5$ (offers gift)	0.3	0.7	0.0
$\alpha 6$ (forces gift)	0.3	0.7	0.0
$\alpha 13(t)$ (waits)	0.05 / sec	0.95 / sec	0.0
$\alpha 14$ (requests payment)	0	0.0	1.0
$\alpha 16$ (concedes gift)	1.0	0	0

Table 4.3 shows the belief mass distribution of various actions affecting `CB(S, Q-Gift, Client, Client)`. $\alpha 5$ is the action of offering the flower as a gift, and it represents a weak evidence towards `Q-Gift` being true. $\alpha 6$ is the action of forcing the flower on the client – this can be interpreted either as an evidence for `Q-Gift`, but also towards its opposite. Both mass distributions keep significant uncertainty. Depending on the belief the agent started from, after these actions the client might still be mostly inclined to believe the flower not to be a gift. Leaving the client with the flower without asking for money (action $\alpha 13$) provides evidence towards the flower being a gift. Every second passing provides more evidence towards `Q-Gift`. Action $\alpha 14$ requesting payment will immediately clarify that the flower is not a gift, and will reduce the uncertainty to zero. In contrast, action $\alpha 16$, conceding the gift, will set the CB to 1.0, also reducing the uncertainty to zero. This action, however, is only a fictional one, which might be expected by an uninformed client, but will never be performed by the seller.
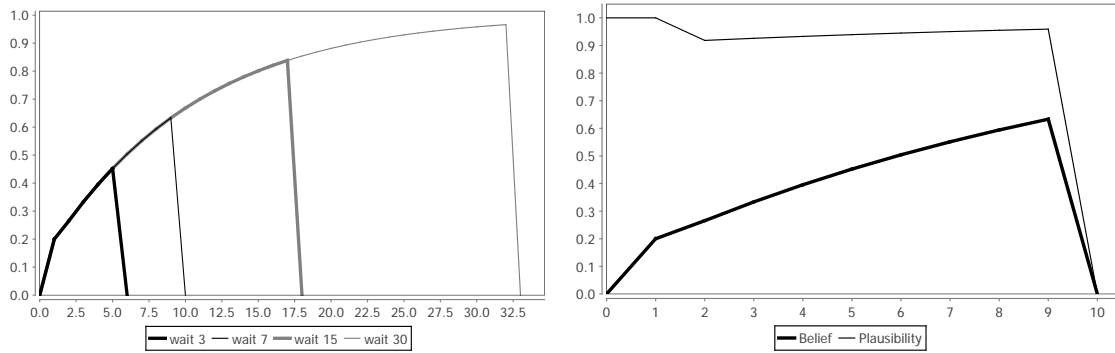
Figure 4.4: The evolution of the value of `CB(S, Q-Gift, Client, Client)` for the action sequence $\alpha5$, $\alpha6$, $\alpha13(t)$, $\alpha14$. Left - the CB value for t = 3,7,15 and 30. Right: the evolution of belief and plausibility for t=7.

Let us now consider an example of how these values fit together in a typical scenario: the client is offered a flower, he refuses, the flower is forced on him, then after a period of wait of 3-30 seconds, he is asked for money. For this scenario we can assume that the client starts with a complete uncertainty (the full belief mass being in $\{T, F\}$). The actions affecting the belief in this case are `CB(S, Q-Gift, Client, Client)` are $\alpha5$ (offer gift), $\alpha6$ (force gift), $\alpha13(t)$ (wait) and $\alpha14$ (ask for money). Figure 4.4-left shows the evolution of the belief for values of t of 3, 7, 15 and 30. Note that the longer the wait, the higher the belief reached. Although only the belief component is used in the CSSM calculations, it is useful to look at the simultaneous evolution of the belief and plausibility (Figure 4.4-right). We find that the incoming evidence not only affects the belief, but also gradually decreases the uncertainty of the client (seen as the gap between belief and plausibility). After action $\alpha14$ not only the

76

belief is zero, but the plausibility as well - thus any further evidence towards `Q-Gift` will be ignored.

4.2   Case studies

In the introduction of this paper we specified that we are requiring two capabilities from our model:

- **Explanatory power:** Let us assume that we have observed a specific instance of the scenario, including its initial conditions, the social agents acting in it, the actions taken, and the outcome of the scenario. We are interested in a consistent explanation of the agent's actions. As a side effect of this explanation, we also hope to gain insight into the decision making process of the social agent.

- **Predictive power:** Let us assume that we are witnessing the initial conditions of a new scenario or, alternatively, we are at an intermediate point in the scenario instance. We want to predict the next action and, by extrapolation, the overall outcome of the scenario.
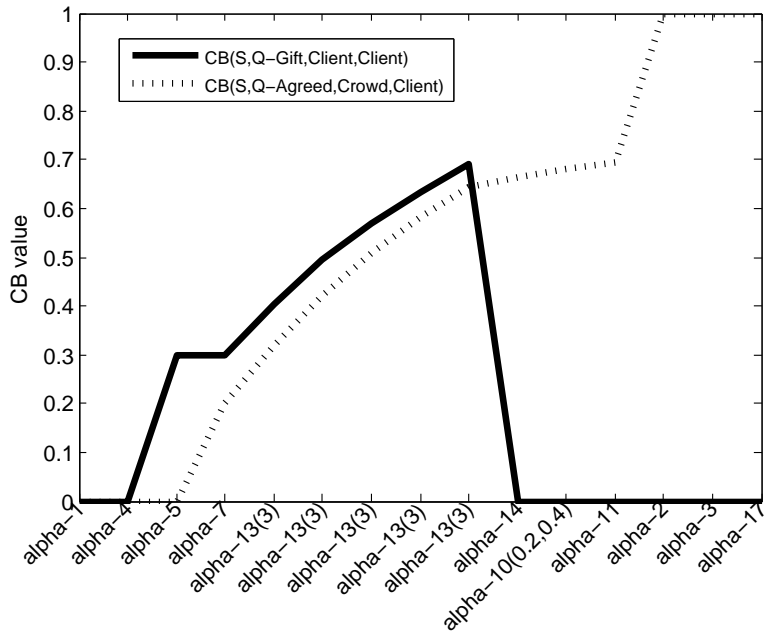
A social calculus model with explanatory and predictive power is not, by itself, sufficient to build a social software agent or robot because such an agent will also have other objectives not captured by social behavior. However, the social calculus model can be an important component of the decision making engine of such an agent.

In the following, we use the model built in the previous section to track two real-world scenario instances witnessed at the Spanish Steps, Rome, Italy.
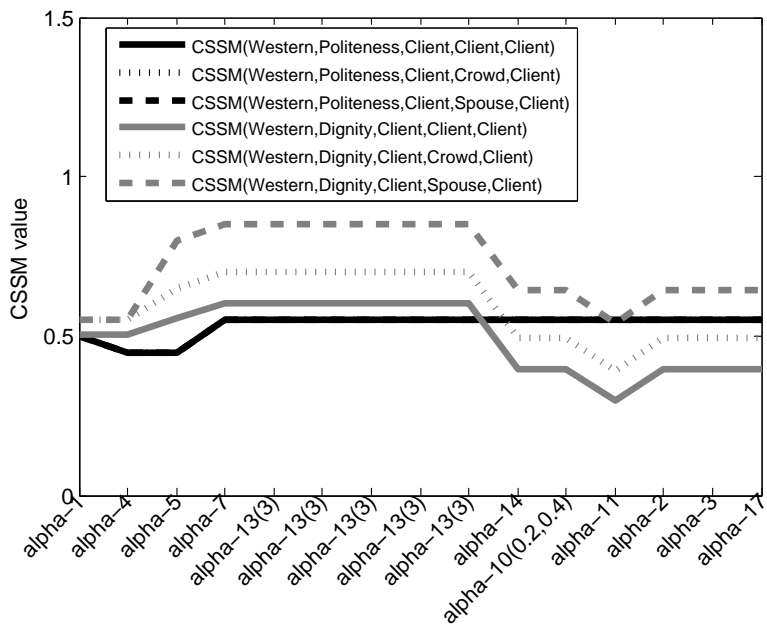
4.2.1 Case 1: Successful sell

In the first observed scenario the seller was successful in selling the flower to a romantic couple. The seller offered the bouquet to the man ($\alpha1$), but was declined ($\alpha4$). Then, the seller offered a flower to the woman ($\alpha5$), and she accepted it. After a waiting time of 15 seconds some distance away ($\alpha13(15)$), the seller returned and requested payment from the man ($\alpha14$). The client attempted to return the flower, with low voice and suggestion type mitigation level ($\alpha10(0.2, 0.4)$). The seller declined to take back the flower ($\alpha11$). At this point, the man accepted to pay ($\alpha2$) and paid for the flower ($\alpha3$).
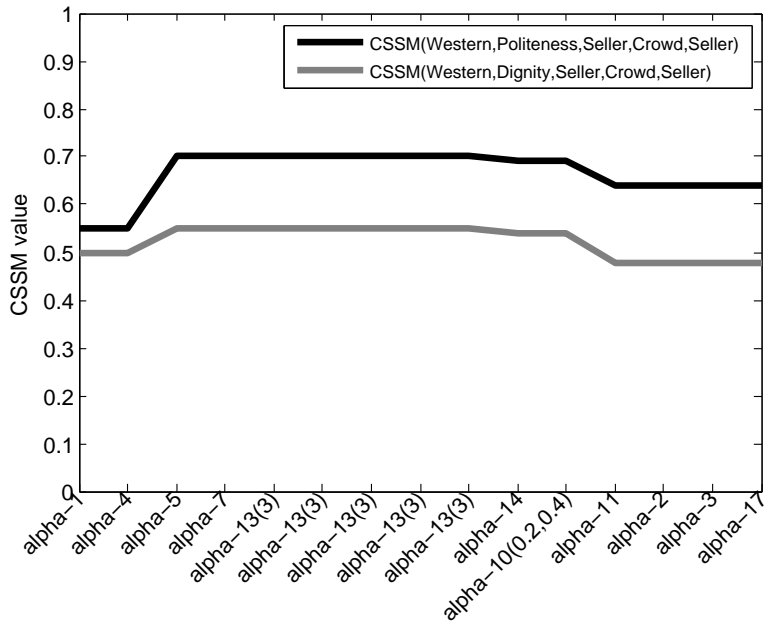
$$\text{TS} \xrightarrow{\alpha1} \text{S1} \xrightarrow{\alpha4} \text{S3} \xrightarrow{\alpha5} \text{S4} \xrightarrow{\alpha7} \text{S7} \xrightarrow{\alpha13(15)} \text{S7} \xrightarrow{\alpha14} \text{S8} \xrightarrow{\alpha10(0.2,0.4)} \text{S9} \xrightarrow{\alpha11} \text{S8} \xrightarrow{\alpha2} \text{S10} \xrightarrow{\alpha3} \text{TP2}$$
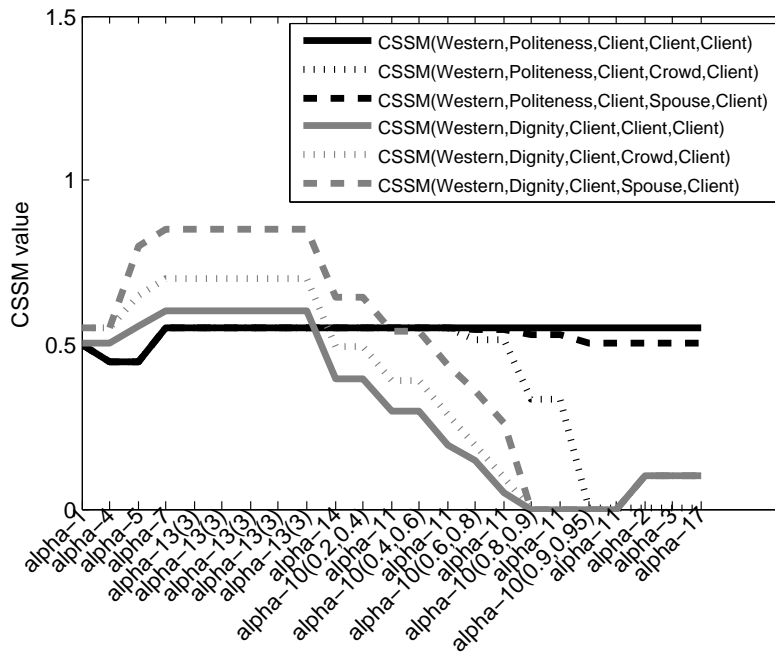
(a) concrete beliefs

(b) client politeness and dignity

(c) seller politeness and dignity

(d) client: what would have happened?

(e) seller: what would have happened?

Figure 4.5: Scenario 1: Successful sell: (a) the evolution of the CBs `Q-Gift` and `Q-Agreed`, (b) the politeness and dignity CSSMs of the client, (c) the politeness and dignity CSSMs of the seller (d) the politeness and dignity CSSM of the client in a fictional "what would have happened?" scenario, e) the politeness and dignity CSSM of the seller in a fictional "what would have happened?" scenario

What requires explanation in this scenario is the fact that the client gives in relatively easily, despite the fact that he does not want the flower (as he tries to return it) and he knows that he is being cheated. Figure 4.5 shows the results of tracking this scenario using our model. For all the graphs, the X axis lists the actions and their parametrization.

Figure 4.5a shows the evolution of the concrete beliefs. The `CB(S, Q-Gift, Client, Client)` starts with a zero value, then it raises to about 0.32 after $\alpha5$ (the offering of a single flower). This appears to be sufficient for the client to accept the flower as a gift. Albeit this value appears to be low, note that this is the Dempster-Shafer belief value which does not imply that the client has a 0.68 belief in the fact that the flower is *not* a gift - the majority of the remainder of the belief mass is concentrated in the uncertainty domain $\{T, F\}$. The belief that the flower is a gift will actually climb during the waiting time of action $\alpha13$ which means that if the client did not give back the flower initially, it will be unlikely that he will give it back during this wait. This statement assumes that no other event changes the client's belief throughout the wait. We have witnessed scenarios where the client holding the flower had seen another client being asked for money, and rushed to return the flower himself, illustrating how actions in one scenario can change CBs in another. This situation can be modeled by our framework but it is beyond the scope of the examples considered in this section. The belief that the flower is a gift plummets to zero once the client is asked for money.

Let use now see the evolution of the client's estimate of the crowd's belief that a transaction had been agreed upon `CB(S, Q-Agreed, Crowd, Client)`. `Q-Agreed` only tracks the existence of an agreement about a transaction – the actual nature of the transaction changes: up to $\alpha14$ the client believes that the transaction had been gift giving, after $\alpha14$, it is clear that the transaction is a commercial one.

From an initial value of 0, this CB jumps to a value of about 0.25 once the client accepts the flower, and gradually increases as long as the client holds the flower. Whether this is a good estimate of the crowd's belief has no relevance to the scenario as long as the crowd is passive. It is not impossible for the crowd to become an active participant in a scenario - people might intervene verbally or call the police. In fact, it is quite likely that the majority of the crowd members did not notice or follow the transaction.

What is relevant from an explanation and prediction point of view is the fact that at the moment when the client is asked for money and makes his attempt to return the flower, this CB has a relatively high value (about 0.68).

Figures 4.5b and 4.5c track the evolution of the dignity and politeness metrics of the client and the seller. Overall, this particular scenario was a very polite interaction, thus we see only moderate changes in the politeness values. The dignity of the client sees somewhat more variation - it initially increases (when the client believes that his spouse is being honored with a gift) and then decreases - when he realizes that he is being cheated. Overall, the client finishes the scenario with quite high dignity and politeness CSSMs. On the other hand, he was obviously cheated and suffered a financial loss.

From an explanatory and predictive perspective, the question is: why did the client accept to pay for the flower? Could we have predicted this outcome? To answer this we can now create a "what would have happened" scenario, where we follow the observed scenario up to a point, and then change it to see what would happen if the client makes a different decision. Figures 4.5d and 4.5e shows the client's and seller's dignity and politeness

in a scenario where, instead of deciding to pay after the first return attempt, the client escalates his return attempts using louder and louder voice and increasingly rude language and gestures. What we see is that this scenario quickly leads to a catastrophic decay of both the dignity and the crowd-perspective politeness of the client while the public politeness and dignity of the seller had been barely impacted. This asymmetry is due to the fact that the client performs these acts in public in front of a crowd which is estimated to believe that what goes on is actually the reneging of an accepted transaction.

The *explanation* for the client paying is that he could continue his return efforts only at a very high cost for his public politeness and dignity while the seller can afford to decline the return of the flower with minimal impact to his public perception.

Similarly, after action $\alpha 13$ our model allows us to *predict* that the selling action will be successful. Note that this prediction is, of course, only probabilistic – human decision making is complex and it is quite possible that some clients will be willing to take hits to the dignity and politeness perception in stride and pursue the return of the flower.

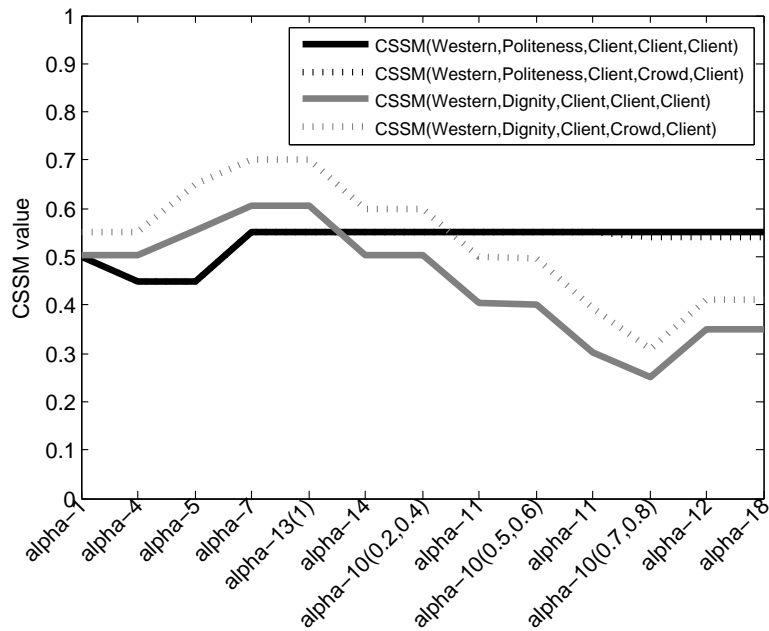## 4.2.2 Case 2: Unsuccessful sell
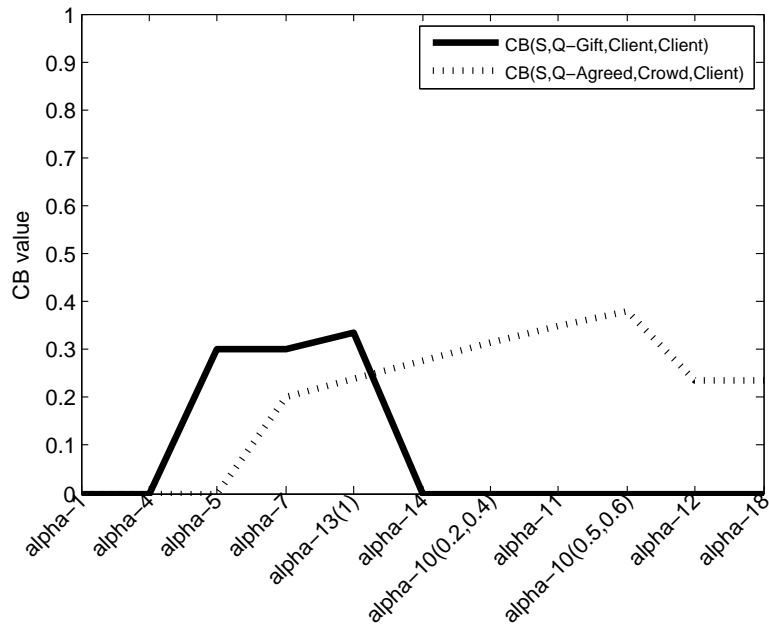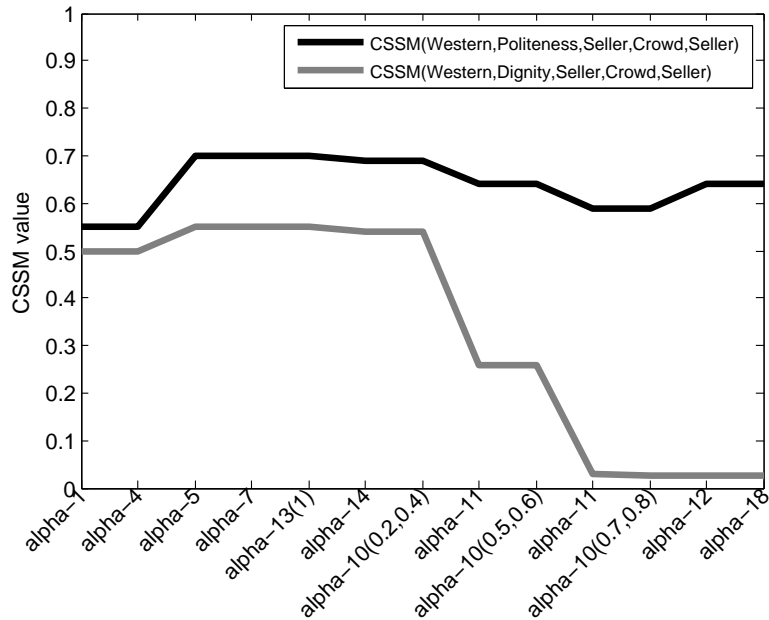

(a) concrete beliefs
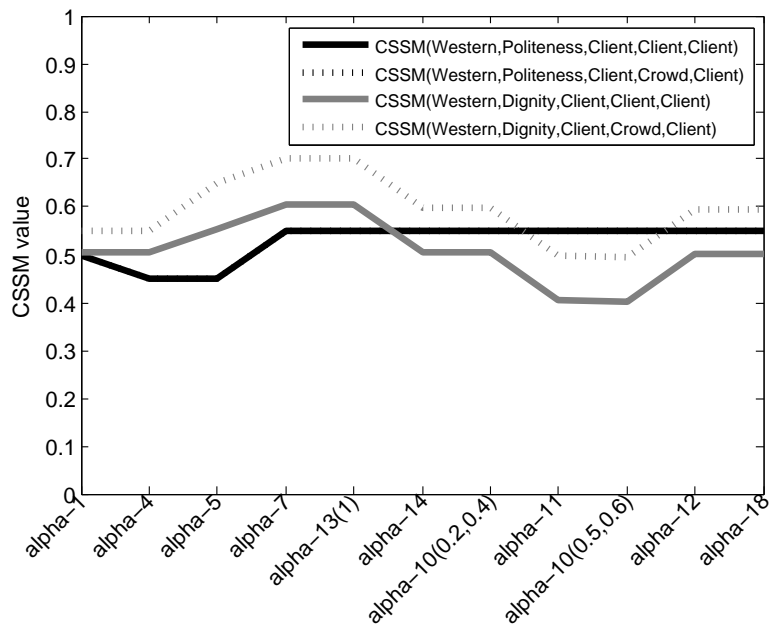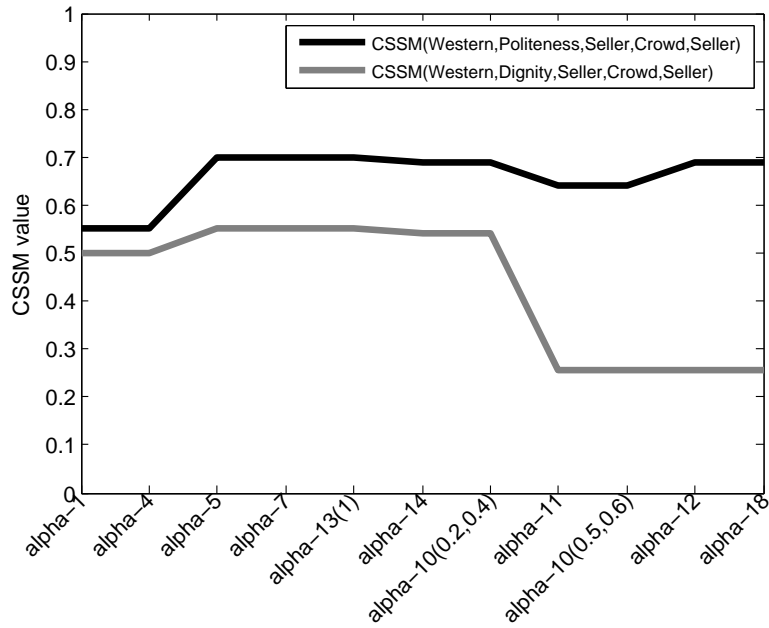
(b) client politeness and dignity

(c) seller politeness and dignity

(d) client: what would have happened?

(e) seller: what would have happened?

Figure 4.6: Scenario 2: Unsuccessful sell: (a) the evolution of the CBs `Q-Gift` and `Q-Agreed`, (b) the politeness and dignity CSSMs of the client, (c) the politeness and dignity CSSMs of the seller (d) the politeness and dignity CSSM of the seller in a fictional "what would have happened?" scenario, (e) the politeness and dignity CSSM of the seller in a fictional "what would have happened?" scenario

The second scenario shows an instance where the seller was unsuccessful in selling the flower. In this case, the client was a single woman. The start of the scenario was similar to the previous case. However, as the woman moved to leave the area, the seller asked her for money only one second after the flower was accepted. The woman had attempted a return in polite terms and low voice $\alpha 10(0.2,0.4)$. After the return was declined, the woman in firm

terms but without using expletives ordered the vendor to take back the flower ($\alpha10(0.5,0.6)$).

At this point the vendor accepted the return ($\alpha12$).

$$\text{TS} \xrightarrow{\alpha1} \text{S1} \xrightarrow{\alpha4} \text{S3} \xrightarrow{\alpha5} \text{S4} \xrightarrow{\alpha7} \text{S7} \xrightarrow{\alpha13(1)} \text{S7} \xrightarrow{\alpha14} \text{S8} \xrightarrow{\alpha10(0.2,0.4)} \text{S9} \xrightarrow{\alpha11} \text{S8} \xrightarrow{\alpha10(0.5,0.6)} \text{S9} \xrightarrow{\alpha12}$$

$$\text{TN2}$$

Figure 4.6a, 4.6b, 4.6c shows the evolution of the CBs, the client's politeness and dignity and the seller's politeness and dignity respectively. To avoid unnecessary repetitions we will concentrate on the differences from the successful sell scenario. The first observation is that the client being a single woman, the spouse-perspective values are not present in the client's evaluation.

In the CBs the main difference is that as the seller was in a rush to ask for money, the asking for money $\alpha14$ happens at a much lower value of `CB(S, Q-Agreed, Crowd, Client)`.

With regards to the CSSMs, both participants end up with a lower dignity. The politeness, however, is relatively unaffected: the client does not use very rude words and gestures (and is protected by the fact that the `Q-Agreed` CB is relatively low). The seller looses some politeness by his first refusal to take back the flower, but recovers in politeness when it accepts the return. He looses relatively large measures of dignity by his refusal. The reason for this is that his estimate of the public belief in an agreement is the same as the client's: `CB(S, Q-Agreed, Crowd, Seller) = CB(S, Q-Agreed, Crowd, Client)`. This means that in his estimate, the crowd is more likely to see this as a forced transaction by a crooked seller, which would make his refusal have a larger impact. Overall, the seller finishes

88

the scenario with acceptable values of dignity and politeness. On the other hand, he did not make a profit and wasted time.

From an explanatory and predictive perspective the question we must ask is why the seller gives up in this particular scenario, and whether it was possible to predict this outcome. Again, we will create a "what would have happened" scenario, where we assume that the seller, instead of giving in, would have repeatedly declined the return (action $\alpha 11$) in the face of more and more insistent return efforts from the client. The client and seller's CSSM's for this hypothetical scenario are shown in Figures 4.6-d and 4.6-e.

What we find in this case is that the seller would suffer a catastrophic decline in public dignity which would not be socially acceptable to him. On the other hand, the client had only a moderate decrease in the dignity and politeness during this escalation phase. This explains why the seller decided to give up the transaction without insisting further.

From a prediction point of view, after $\alpha 13$ and $\alpha 14$, that is after the seller asks for money prematurely, we can predict that the sell will likely fail, because the client can push the seller into deep public dignity loss while suffering relatively minor damage to her own politeness and dignity. Again, this prediction is probabilistic and depends on the willingness of the client to start the escalation of the return effort - if the client gives up after the first try, the scam can still succeed.

CHAPTER 5
SOCIAL CALCULUS - LEARNING SOCIAL BEHAVIOR

Social skills, the ability of an individual to manage social relationships and cultural values to achieve its own goals is a profitable human skill, not necessarily associated with high moral values nor intelligence in the abstract sense. Con artists and crooked merchants must have significant social skills. On the other hand many highly intelligent individuals exhibit Asperger's syndrome which associates with reduced ability to navigate social interactions. There is a widespread belief that the ability to navigate the complexities of social and cultural interactions is one of the most difficult tasks for artificial intelligence entities (agents or robots) to achieve.

The objective of this chapter is to investigate the degree to which strategies associated with successful manipulation of a social scenario are learnable. To investigate this in the Spanish Steps flower selling scam, a crooked seller tries to exploit the self-evaluated values of politeness and dignity, as well as the peer and public image of a prospective customer to sell him overpriced flowers. The clients are aware that they are being cheated: the fact that the scam sometimes succeeds requires a very precise manipulation of the social sentiments by the crook. In the real world, both the seller and the client learn from this scenario, but while the seller has many opportunities to learn a social behavior strategy which leads to an occasional success, a typical victim of the scam had usually participated in the scenario

the first time. In addition, most of the victims are foreign tourists who are unsure about their estimates for the social metrics in the new environment. Our goal is to reproduce the process through which an apprentice crook can improve its social manipulation strategies. We can state our goal as follows: learn a strategy where the Seller actor, whenever it is in its turn to take an action, will take actions which manipulate the CSSMs and CBs involved in such a way as to lead the scenario to a progress state desired by the seller (which involves the buyer buying the flower).

5.1 Training and Execution Algorithm

In the following, we will describe a technique through which the seller agent can improve its strategy through repeated interactions with different client agents. The overall technique is one of reinforcement learning – successful sells reinforce the behaviors which led to their success. Note that the seller agent does not learn the whole scenario from scratch - we assume that the general flow of the scenario is a given and had been communicated to the apprentice seller. Similarly, the seller already has certain behavioral rules - for instance it will not pursue the sell to a catastrophic loss of his own public dignity. What the seller actually learns is an ability to fine-tune his selling ability.

In order to train the seller, we introduced a number of consistent policies for the clients to generalize the scenario. Consistent policies meant to provide consistent behavior of the client in the scenario. The consistent behavior of clients help seller agent to learn the

social matrix, evaluate and identify appropriate actions in response. We are not considering any cooperative or adversarial game strategy among the agents. Clients traverse from one state to another enacting neutral behaviors: they don't get any rewards for cooperation neither gets penalized. In this consistent policy, the agent takes the default action in the state where only one choice is available. For example, choosing α_1 in the state $S0$, where seller at the start of interaction can only offer flower or choosing α_3 in state $S10$ where the client has to pay the price after accepting the deal in public. The agent also carries an acceptable lower bound over social-cultural matrix values. For example, the seller can refuse to accept the flower in the state $S9$ by considering the action α_{11} till the client lowers its social politeness by yelling loud in an offensive manner, which in effect brings down the seller's self-dignity and its dignity in crowd's perspective. Consistent policy is constant in terms of lower bound of CSSMs, but not in terms of state - action pairs. For example, the seller occasionally, gives up, without pushing the selling process to its extremes. This is because either the interaction interval has stretched too long or either the crowd has turned arrogant due to excessive and frequent deceptions. These five consistent policy agents for the clients not only help in generalizing our model but also effectively cross-validate our model against different behavior models. The main idea to pose a number of different clients is to train the crook-seller well enough to performing the deceptive sale in presence of crowd while maintaining his CSSM values. These consistent policy behavior of the clients that we introduce in our experimental model are as follows: *casual, busy, arrogant, smart and wealthy.*

- The *casual client* is easy to handle, not offensive in speech, has appropriate loudness factor and is afraid of losing face value in the public. This client proves an easy prey for crook natured seller.

- The *busy client* has no time to spare on conflicts and would like to end the conversation the earliest. If the crook-seller manages to notice this client's behavior, then he will play the deception tactics with higher probability. The busy client tends to buy the flower offered to him after long conflicting arguments to end the conversation.

- The *arrogant clients* are the ones who do not care for the face values in public. Being offensive in nature provides high values for offensiveness and loudness variables. They are the ones who ready to throw the gifted-flower on the floor, even though the seller was bent upon posing it as a gift for the client.

- *Smart clients* are the wise ones and keenly observe the crook sellers' perception. They will maintain public dignity and politeness and still are able to convince the seller to take his offer of gift back.

- The *wealthy clients* readily purchase the flower from the seller based on random variable which takes on the values of a flip-coin.

Q-learning based reinforcement learning generates a map of state-action pairs based on real numbers called Q-values. The Q-values help the agent in making a decision about the optimal action in a given state. We are using function approximation based Q-learning where we are defining Q-values as a linear combination of weighted features. These features

are CSSMs in the framework. In the Spanish Step scenario, the progress based state-space model which was a combination of state-action pair of all agents together instead of single action state space representation of an individual agent. In order to apply Q-learning method we need to draw-out individual agent's progress function state space. We consider feature based Q-function as:

$$Q(s, a) = \sum_i w_i f_i(s) \tag{5.1}$$

The Q-value is not dependent upon the action because the extracted features depend upon the state only. For an agent in state s, its moves into state s' by taking action a, the reward is defined as $r(s, a, s')$ and is updated with the new feature's values from Action-Impact function. Now we need to calculate a new estimate of the Q - value for s based on this iteration which is given as

$$Q_{new}(s, a) = r(s, a, s') + \gamma \cdot \sum_i w_i f_i' \tag{5.2}$$

where gamma is a discount factor on future rewards. The feature weights get updated with this new estimated as follows

$$w_i = w_i + \alpha \cdot (Q_{new}(s, a) - Q_{old}(s, a)) \cdot f_i(s) \tag{5.3}$$

where alpha is a learning rate for weight update. The advantage of using function approximation based Q-learning is that we need not to remember the Q-Value for ever state-action pair, and also for features based on the states.

Algorithm 1 and Algorithm 2 show learning process and off-policy based function approximation Reinforcement learning pseudo code. Both methods use linear approximation

Algorithm 1 Approximation function based learning

1: Initialize $\vec{w}$ arbitrarily

2: **while** No client left to interact **do**

3:     $s, a \leftarrow$ Initial state and action

4:     $\mathcal{F}_a \leftarrow CSSMs$

5:     **while** s' is terminal state **do**

6:        $r \leftarrow observethereward$

7:        For all $i \in \mathcal{F}_a : Q_{old} \leftarrow \sum_{i \in \mathcal{F}_a} w_i \mathcal{F}_a(i)$

8:        For all $i \in \mathcal{F}'_a : Q_{new} \leftarrow \sum_{i \in \mathcal{F}'_a} w_i \mathcal{F}'_a(i)$

9:        For all $i \in \mathcal{F}_a : w_i \leftarrow w_i + \alpha \cdot (r + \gamma \cdot Q_{new} - Q_{old}) \cdot \mathcal{F}_a(i)$

10:     **end while**

11: **end while**

Algorithm 2 Approximation function based execution

1: Initialize $\vec{w}$ with learned weights

2: **while** s' is terminal state **do**

3:     update(w)

4:     $s, a \leftarrow$ Initial state and action

5:     **if** $random(p) > 1 - \epsilon$ **then**

6:        **for** $a' \in \mathcal{A}(s)$ **do**

7:           $\mathcal{F}_{a'} \leftarrow$ set of CSSMs present in s', a'

8:           $Q_{a'} \leftarrow \sum_{i \in \mathcal{F}_{a'}} w_i \cdot \mathcal{F}_{a'}(i)$

9:        **end for**

10:       $a \leftarrow argmax_{a'} Q_{a'}$

11:     **else**

12:        $a \leftarrow$ a random action $\in \mathcal{A}(s)$

13:     **end if**

14: **end while**

with CSSM features normalized between [0, 1]. In the turn-based scenario such as Spanish Step, the agent can not visualize the reward from immediate next state. Therefore, weights have been updated backward. The set of feature $\mathcal{F}_{a'}$, corresponds to the current state and all possible actions a'. $\mathcal{F}_a$, corresponds to the previous state and action a applied to achieve current state. Off-policy method uses an ϵ-greedy policy for choosing next action.

5.2   Experiments

The experiment phase is divided into three phases. The initial phase is the formalization of the ActionImpact function as shown in the Chapter 4. Next comes in the implementation of the reward values for the reinforcement learning. The last phase is the implementation of the function approximation based reinforcement learning.

Table 5.1: Reward Value Table

(s, α, s')	Reward	Description
$(S5, \alpha9, TN1)$	-100	Client throws the flower to the ground
$(S2, \alpha3, TP1)$	0	Client pay for the flower
$(S3, \alpha15, TN1)$	-100	Seller quit before offering flower as gift
$(S6, \alpha12, TN1)$	-100, 100	Seller decline to take flower back
$(S7, \alpha16, TF2)$	-100	Seller concede the gift
$(S9, \alpha12, TN2)$	-100, 100	Seller decline to take flower back
$(S10, \alpha3, TP2)$	-100, 100	Client pay for the flower after long effort

We will describe the reward-values as shown in the Table 5.1. In Table 5.1, the first column is the state-action-state transition represented as (s, α, s'). The second column defines the reward value for the state-action-state transition. The third column describes that transition of states using the actions. $(S5, \alpha9, TN1)$ represents transition for the client throwing the flower on the ground. The seller has to avoid this transition which can not only dishonor his dignity but would also cause him loss of money, hence the reward is negative value. $(S2, \alpha3, TP1)$ represents transition for the client when he pays for the flower. In this transition, the seller has no contribution and its totally dependent on the client type. Therefore, no reward been earned during this transition. $(S3, \alpha15, TN1)$ represents transition where the seller choose to quit before offering flower as gift. This transition is the naive form of selling where seller moves from one client to another without pressurizing

either. $(S6, \alpha12, TN1)$ represents the transition where the seller had decided to give up and moved to the next client.

The seller rewards for the time used for successful transactions, otherwise rewards negative for wasting his time with the wrong client. $(S7, \alpha16, TF2)$ represents transition where the seller decided to concede the flower that was given as gift. This is most un-likely state for the seller to perform because the seller will lose money in this transition. $(S10, \alpha3, TP2)$ represents transition where seller will be rewarded for insisting the client for sufficient time duration before quitting the game.

5.2.1 Results
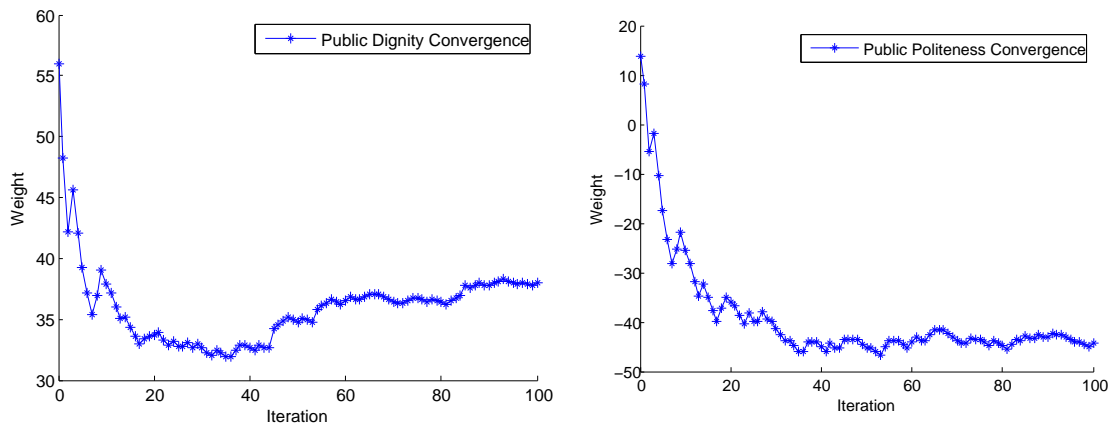
Figure 5.1: Seller Public Politeness and Public Dignity Weights Convergence
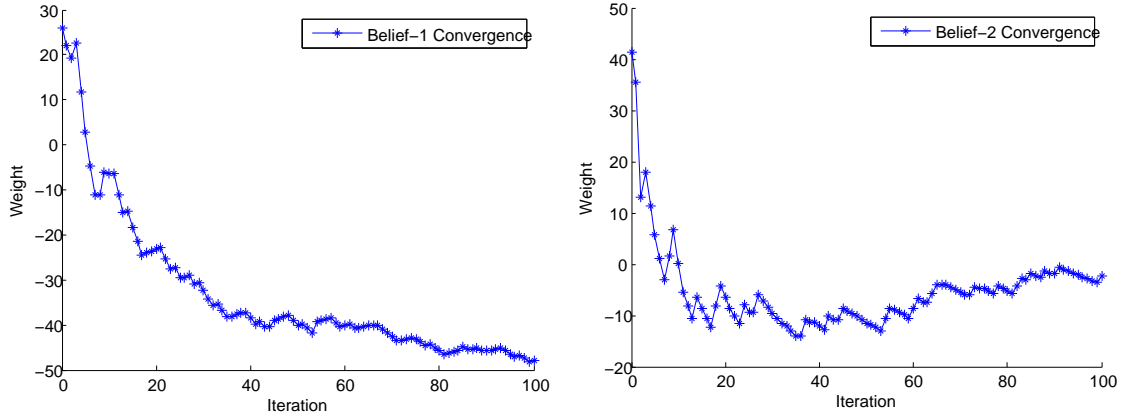
Figure 5.2: Seller Belief-1 and Belief-2 Weights Convergence

We used a scripting based simulator to generate state transitions, observation and immediate rewards for interaction of the seller with five different consistent policy behavior agents. Each time the conversation reaches the terminal state, it is reset to the initial configuration. We allowed the simulator to learn weights for the population using 500 iterations where each iteration is equivalent to 200 time steps. This ensures sufficient learning over all of the possible modeled state transitions. The learning rate decreases with the inverse factor of the iterations. Convergence of weights of CSSMs and CBs acting as a features of the algorithm are shown in the Figure 5.1 and 5.2. Figure 5.1 plots the feature weight learning for public politeness and public dignity of the seller as perceived by crowd from the seller's estimation. Belief-1 also known as is_a_gift is the individual concrete beliefs of the seller. This belief helps the seller in posing the fraudulent transaction that the given flower was gift. Belief-2

also known as *has_been_a_transaction* helps the seller to gain the attention from the crowd to make his action more legit for selling his flowers.

CHAPTER 6
SOCIAL CALCULUS - BELIEF AND SOCIAL METRIC PROPAGATION

The simulations where CSSMs had been deployed, however, up to this point were always considering a single interaction of several minutes at a time. However, the public perception can evolve over longer time frames spanning multiple interactions. One of the most intriguing aspects of public perception modeling is the way in which knowledge of individual actions propagates in space and time, how interactions at different spatiotemporal locations affect each other through the public perceptions and how does the general public (such as a crowd of bystanders) forms and forgets a public perception.

The work described in this paper extends the CSSM model towards allowing these type of inferences. For a concrete example, we will use an extended version of the Spanish Steps scenario which follows the interaction of a seller with multiple clients over a longer period of time. We make an effort to realistically model the public perception as provided by the ever changing crowd at a tourist attraction. In this chapter, we discuss the mechanisms for multitasking from the point of view of the seller: how does the perception work? How can these set up next to each other? How does the knowledge and beliefs propagate among the simultaneous clients of the same seller?

6.1 Multitasking

6.1.1 The action-state graph

The unfolding of the Spanish Steps scam can be relatively well separated in discrete steps, allowing us to draw an action-state graph as shown in Figure 6.1. This graph is not a full description of the interaction, only an aid in organizing our representations. Being in a certain node does not fully represent the state of the scenario - we need also to consider a number of *detail variables*. For instance, S6 is a state where the client holds the flower and had just attempted to return it to the seller. The details of this state include the judgment by the seller and the client of the current situation, as well as their emotional state. If the client believes that the public assumes that she had already accepted the transaction, she will be more reluctant to force the return.

Similarly, the actions represented by the edges of the graph are also parameterized by detail variables. In our model, A7, A9, and A16 are parameterized by their "loudness" x which determines how many onlookers will overhear the transaction and their "offensiveness" y which will determine how the action will impact the values of the actor and target of the action. The action A14 is parameterized with the waiting time t it involves.
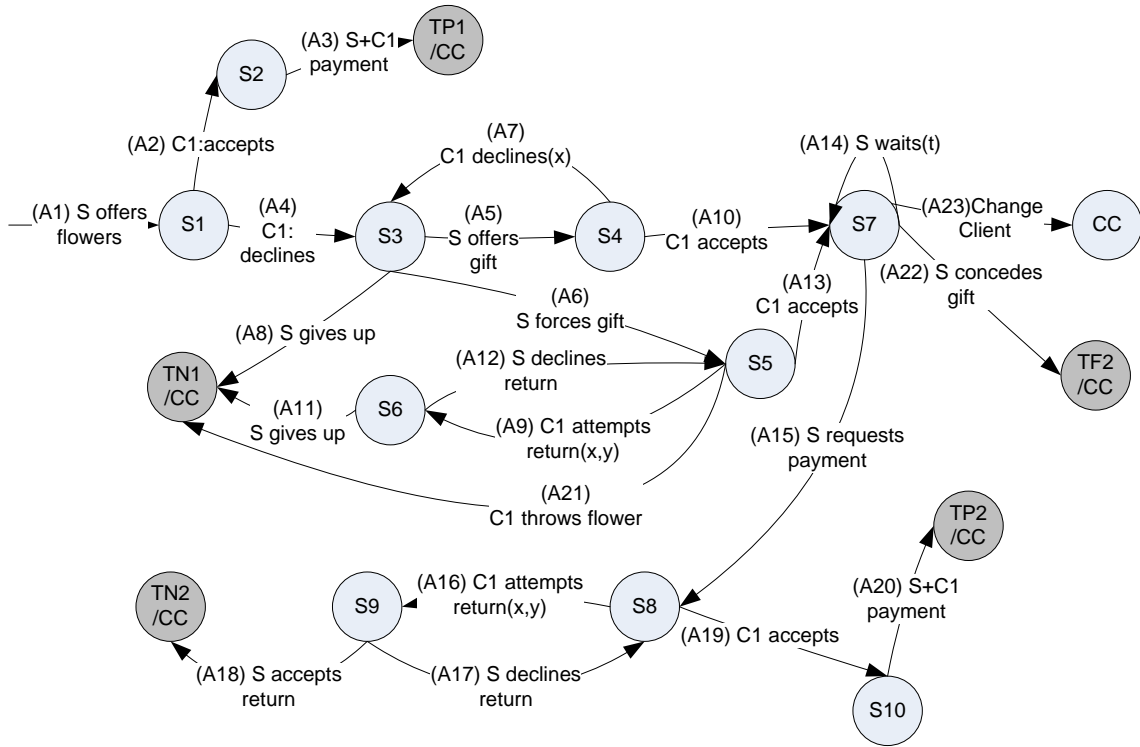
Figure 6.1: The progress graph of the Spanish Steps scenario. The states marked with CC allow for the change of clients.

The seller in the Spanish Steps scam can not execute more than one action at a time, even if it involves multiple clients. Furthermore, basic rules of social interaction, such as the necessity to maintain physical proximity and eye contact prevent the seller from arbitrarily switching between clients. However, the Spanish Steps scam has certain states where switching away from a client *is* possible, and in some cases, such as state S7, even desirable. Exploiting these states, the seller can handle multiple simultaneous transactions, each in a specific state.

As the seller interacts physically with the clients, the clients will necessarily be in close physical proximity, and they will also likely be paying attention to the seller. Thus, we can make the assumption that the events unfolding in the parallel threads will be known to all the participants, and influence their beliefs.

To model the actions of the seller, we have designated some of the states in the state-action graph in Figure 6.1 as *change client* (CC) states. These are states where the seller has the possibility to either start a new interaction, by approaching a new client, or to resume the interaction with an existing client. Naturally, all the terminal states of the graph are CC states - in this case the interaction is terminated and the seller does not need to return to the client. State S7/CC, however, is not a terminal state: the seller will need to return to the client holding the flower.
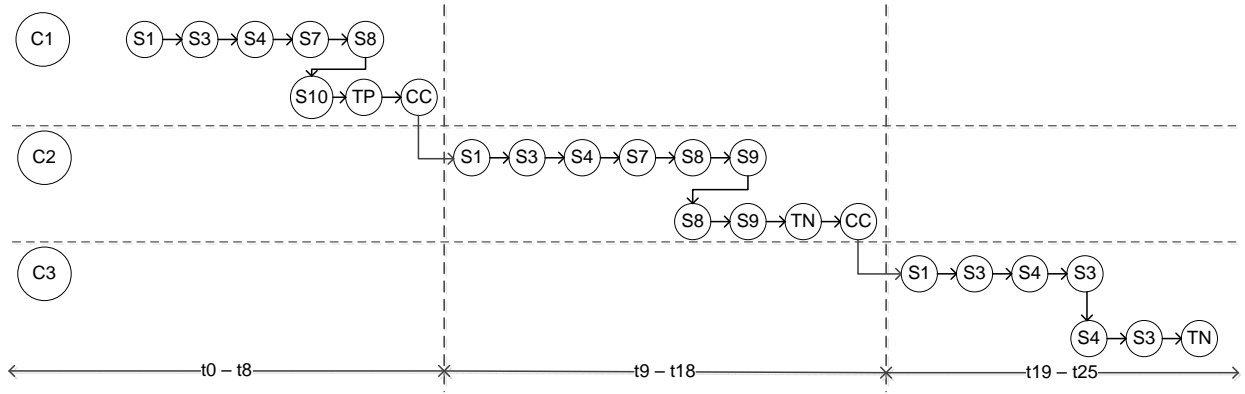
(a) serial interaction
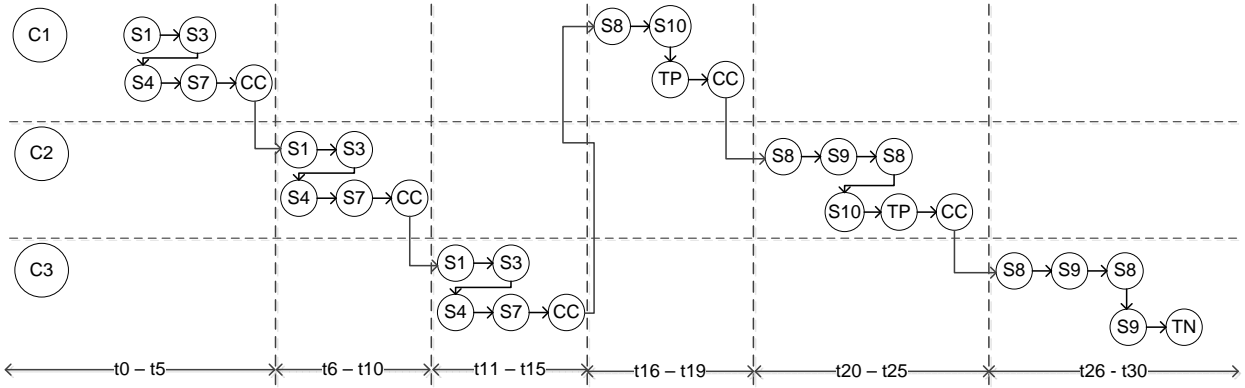
(b) interleaved interaction

Figure 6.2: Two possibilities for handling multiple clients in the Spanish Steps scenario

Fig. 6.2a shows the flow of three instances of the scenario where transitions are only made at terminal states. We call this a *serial interaction*. A serial interaction is not equivalent to three separate scenarios. While there is no overlap between the scenarios, there is a leak of information from one scenario to the next. This happens through two mechanisms: (a) through the clients in the later scenarios directly witnessing the outcomes of the previous scenarios, and (b) through the impact of the scenarios on the public perception.

105

Fig. 6.2b shows an example where the seller *interleaves* the interaction with three different clients. In this case, the close physical proximity guarantees that the clients are aware of the unfolding of the scenario with the other clients. One would think that more information would help the clients, but this is not necessarily the case: the received information can actually be deceptive. The seller can actually derive an advantage from multitasking, beyond the purely time saving aspect. Let us consider the case of client C3 when entering the scenario, at state S1. For the sake of simplicity, let us consider that C3 had witnessed the evolution of the scenario of C1 and C2. In the scenario described in Fig. 6.2a, C3 had seen the complete unrolling of the scenario two times. She knows that the single flower offered is not a gift, as she had seen the seller ask money for it on two different occasions. Thus C3, although she might choose to buy a bouquet of flowers, if she feels like it, will not fall for the scam, by not accepting the single flower from the seller. Her best choice is to take the path $S4 \xrightarrow{A7(10)} S3 \xrightarrow{A8} TN1$ out of the scenario.

In the scenario described in Fig. 6.2b however, what C3 had seen is that the clients C1 and C2 accepted the single flower and had not been asked for money. This information would encourage C3 to accept the flower, and reach state S7 in the scenario. Note that the client will still be able to escape without paying by escalating the return efforts on the path of the repeated iterations of $S8 \xrightarrow{A16(x)} S9 \xrightarrow{A17} S8$ with increasing values of the parameter x. However, this will be vastly more expensive in terms of time, dignity and politeness.

If the seller does not interleave the clients, his best choice is to pause between the instances for a sufficiently long time such that the client C3 would not have witnessed the

previous scenario. Alternatively, the seller might choose a client who had recently arrived to the scene. One way to achieve this is to move to a different location, to make sure that the bystanders have not witnessed the previous scenario.

6.1.2   Culture-sanctioned social metrics

To model the Spanish Steps scenario we used two concrete metrics: the *financial worth* W and the *time T* and two CSSMs: the *dignity D* and the *politeness P*. Both sides consider the metrics from the perspective of the self and the public; the client also considers a peer (the other member of the romantic couple). With these assumptions, the vector of metrics for the client is $\{W^c, T^c, D^c, D_p^c, D_r^c, P^c, P_p^c, P_r^c\}$ while the vector of the seller is $\{W^s, T^s, D^s, D_p^s, P^s, P_p^s\}$.

6.1.3   Beliefs and public perceptions

Every action of an actor impacts the metrics of his own and his interaction partner. The change in a specific metric, by a specific action, in specific circumstances is given by the *action impact function* (AIF). Let us now investigate mathematical form of AIF. In the first approximation, the AIF depends on the detail parameters of the action. Let us consider action A16 (client attempts return), which is characterized by the *loudness x* and *offensiveness*

y. Obviously, the higher these values, the stronger the effect on the dignity of the seller and the politeness and dignity of the client.

However, the impact also depends on the *beliefs of the public perception* of the scene. For a given level of loudness and offensiveness, it is less of a loss of dignity to be offensive with a crooked merchant than with an honest one. Similarly, one looses more dignity when reneging an agreed-upon transaction compared to correcting a misunderstanding. To model observed behavior of the real world players in the Spanish Steps scenario, we need to consider at least the following beliefs:

B^c_{gift} the client's belief that seller intends the flower to be a gift

B^c_{agr} and B^s_{agr} the client's and, respectively, sellers belief that the general public thinks that a transaction had been agreed upon.

B^{sc}_{agr} the sellers estimate of B^c_{agr}

We consider a number of other beliefs in the scenario involving the periodic interaction of seller over longer span of time. These beliefs include

- B^c_{dec} the client's belief that the seller is deceptive, being a function of past experiences.

- B^w_{dec} the client's belief that the crowd perceives the seller as deceptive, dependent upon the visual or verbal communication with other agents in the crowd and by the cultural understanding of the place

Naturally, beliefs are not orthogonal: a certain action can be evidence or counter-evidence against more than one belief. Furthermore, the way in which beliefs propagate between the agents depend on many factors, including the temporal and spatial aspects of

the scenario. Clients who are in close proximity have a higher probability of information sharing. A tourist who had spent some time in the location has a better knowledge about the seller's deception than a newly arrived crowd member.

6.2 Experimental study

Table 6.1: Experiments

Clients	Actions	Transaction
Experiment 1: Non-interleaving without breaks		
C1	$\xrightarrow[t0]{A1}$ S1 $\xrightarrow[t1]{A4}$ S3 $\xrightarrow[t2]{A5}$ S4 $\xrightarrow[t3]{A10}$ S7 $\xrightarrow[t4]{A15}$ S8 $\xrightarrow[t5]{A19}$ S10 $\xrightarrow[t6]{A20}$ TP2 $\xrightarrow[t7]{A24}$ CC	pass
C2	$\xrightarrow[t8]{A1}$ S1 $\xrightarrow[t9]{A4}$ S3 $\xrightarrow[t10]{A5}$ S4 $\xrightarrow[t11]{A10}$ S7 $\xrightarrow[t12]{A15}$ S8 $\xrightarrow[t13]{A16(0.2,0.2)}$ S9 $\xrightarrow[t14]{A17}$ S8 $\xrightarrow[t15]{A16(0.4,0.4)}$ S9 $\xrightarrow[t16]{A18}$ TN2 $\xrightarrow[t17]{A24}$ CC	fail
C3	$\xrightarrow[t18]{A1}$ S1 $\xrightarrow[t19]{A4}$ S3 $\xrightarrow[t20]{A5}$ S4 $\xrightarrow[t21]{A7(0.6,0.3)}$ S3 $\xrightarrow[t22]{A6}$ S5 $\xrightarrow[t23]{A9(0.5,0.5)}$ S6 $\xrightarrow[t24]{A11}$ TN1	fail
Experiment 2: Non-interleaving with breaks		
C1	$\xrightarrow[t0]{A1}$ S1 $\xrightarrow[t1]{A4}$ S3 $\xrightarrow[t2]{A5}$ S4 $\xrightarrow[t3]{A10}$ S7 $\xrightarrow[t4]{A15}$ S8 $\xrightarrow[t5]{A19}$ S10 $\xrightarrow[t6]{A20}$ TP2 $\xrightarrow[t7]{\mathbf{A14(20)}}$ TP2 $\xrightarrow[t8]{A24}$ CC	pass
C2	$\xrightarrow[t9]{A1}$ S1 $\xrightarrow[t10]{A4}$ S3 $\xrightarrow[t11]{A5}$ S4 $\xrightarrow[t12]{A10}$ S7 $\xrightarrow[t13]{A15}$ S8 $\xrightarrow[t14]{A16(0.1,0.1)}$ S9 $\xrightarrow[t15]{A17}$ S8 $\xrightarrow[t16]{A19}$ S10 $\xrightarrow[t17]{A20}$ TP2 $\xrightarrow[t18]{\mathbf{A14(30)}}$ TP2 $\xrightarrow[t19]{A24}$ CC	pass
C3	$\xrightarrow[t20]{A1}$ S1 $\xrightarrow[t21]{A4}$ S3 $\xrightarrow[t22]{A5}$ S4 $\xrightarrow[t23]{A10}$ S7 $\xrightarrow[t24]{A15}$ S8 $\xrightarrow[t25]{A19}$ S10 $\xrightarrow[t26]{A20}$ TP2	pass
Experiment 3: Interleaved		
C1	$\xrightarrow[t0]{A1}$ S1 $\xrightarrow[t1]{A4}$ S3 $\xrightarrow[t2]{A5}$ S4 $\xrightarrow[t3]{A10}$ S7 $\xrightarrow[t4]{A24}$ CC	hold
C2	$\xrightarrow[t5]{A1}$ S1 $\xrightarrow[t6]{A4}$ S3 $\xrightarrow[t7]{A5}$ S4 $\xrightarrow[t8]{A10}$ S7 $\xrightarrow[t9]{A24}$ CC	hold
C3	$\xrightarrow[t10]{A1}$ S1 $\xrightarrow[t11]{A4}$ S3 $\xrightarrow[t12]{A5}$ S4 $\xrightarrow[t13]{A10}$ S7 $\xrightarrow[t14]{A24}$ CC	hold
C1	$\xrightarrow[t15]{A15}$ S8 $\xrightarrow[t16]{A19}$ S10 $\xrightarrow[t17]{A20}$ TP2 $\xrightarrow[t18]{A24}$ CC	revisited/pass
C2	$\xrightarrow[t19]{A15}$ S8 $\xrightarrow[t20]{A16(0.3,0.3)}$ S9 $\xrightarrow[t21]{A17}$ S8 $\xrightarrow[t22]{A19}$ S10 $\xrightarrow[t23]{A20}$ TP2 $\xrightarrow[t24]{A24}$ CC	revisited/pass
C3	$\xrightarrow[t25]{A15}$ S8 $\xrightarrow[t26]{A16(0.3,0.3)}$ S9 $\xrightarrow[t27]{A17}$ S8 $\xrightarrow[t28]{A16(0.3,0.3)}$ S9 $\xrightarrow[t29]{A18}$ TN2	revisited/fail

In the following we will describe a series of experiments which model the propagation of the public perception across multiple instances of the Spanish Steps scenario.

We have traced the model in three different scenarios. Each of them represent the activities of a seller enacting the Spanish Steps scam with three different clients C1, C2 and C3. The three experiments are described in Table 6.1.

Experiment 1 is an example of a Non-interleaving interaction with no breaks between the scenarios. As soon as the seller finishes a scenario, he immediately chooses the next client and starts the next scenario. Experiment 2 is a Non-interleaving interaction with breaks (delays) between the scenarios. To model the effect of the break, we have applied the Ebbinghaus forgetting curve to all the beliefs of the agents (essentially pulling the Dempster-Shafer values towards ignorance).

6.2.1 B_{gift} and D_p^s

In Experiment 1 the seller was successful with the first client, as he succeeded to raise B_{gift} from 0.5 to 0.8. The second and third clients, however, had witnessed this interaction, thus their own B_{gift} values had started from much lower values. In the case of C3, for instance, the B_{gift} value starts at 0.3. This is so low that it allows the client to reject the offered single flower with high loudness and offensiveness values, which terminates the interaction (unsuccessfully for the seller) at state TN1. Fig. 6.3a and Fig. 6.3b show the evolution of B_{gift} and the seller's public dignity D_p^s for Experiment 1.
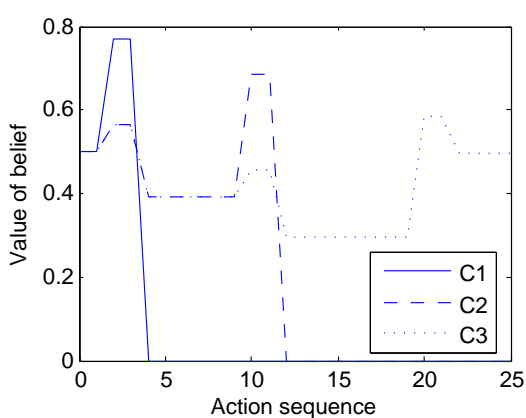
(a) B_{gift} of the clients
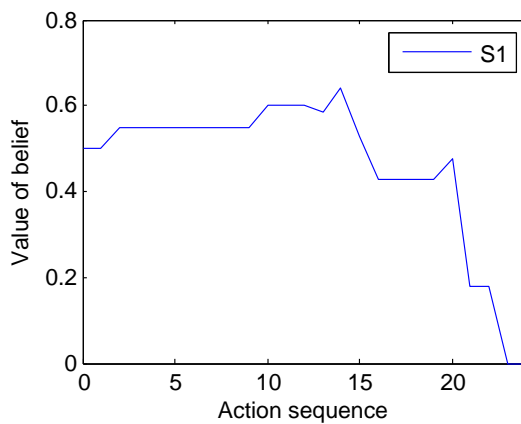
(b) D_p^s of the seller
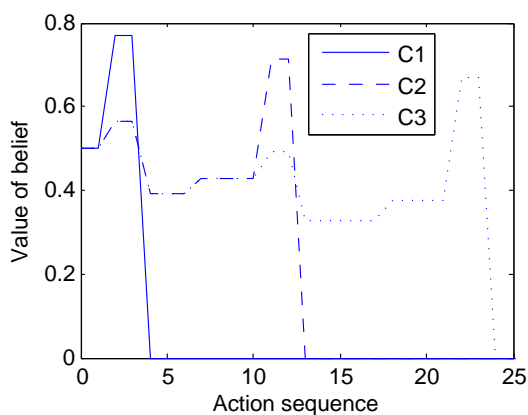
(c) B_{gift} of the clients
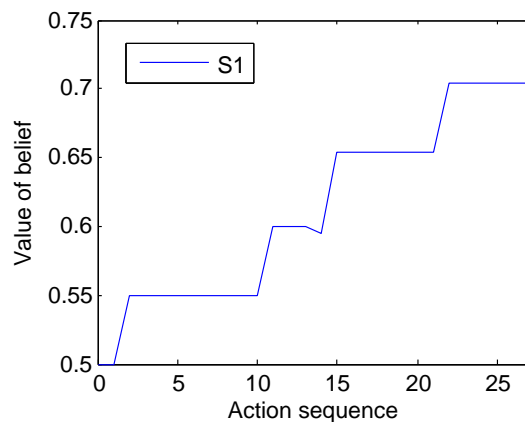
(d) D_p^s of the seller
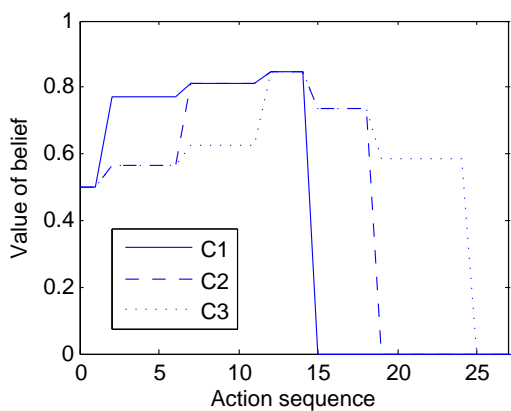
(e) B_{gift} of the clients
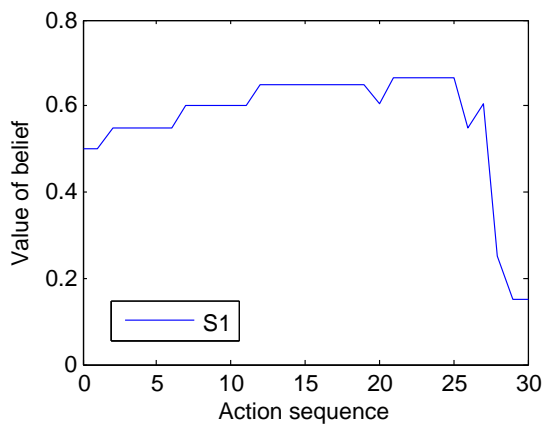
(f) D_p^s of the seller

Figure 6.3: Non-interleaving without breaks (top row), Non-interleaving with breaks (middle row), Interleaving clients (bottom row)   112

In the second experiment, the seller performs the same scam, but this time he takes a break between the individual clients. This break guarantees that the clients did not see the unfolding of the previous scenarios, and the public perception had also returned to neutral. This is a result of both the gradual turnover of people in the crowd of the tourist attraction, and the natural forgetting of the individuals. As a result, all the clients are essentially starting from a neutral point. In Experiment 2 the seller had involved three clients in the scam successfully. Naturally, we can have instances where a client would be able to avoid the scam in this case as well, by escalating the loudness and offensiveness of her return efforts. However, even if she avoids the scam, the client will loose significant amount of dignity and politeness CSSMs, because she does not have the favorable support of the public. Fig. 6.3c and Fig. 6.3d show the evolution of B_{gift} and D_p^s for Experiment 2. Note, however, that taking long breaks is not an efficient way for the seller to maximize his profit W^s.

Experiment 3 shows an example of interleaved scenario. In this case, the clients are in close proximity, and aware of each other. However, up to state S7 neither they, nor the general public will be aware of the full flow of the scenario, thus they will actually have a higher B_{gift} then the two previous cases. On the other hand, once the seller starts to ask the clients for money, this information is quickly propagated to the remaining clients and the public perception as well. As a result, the public perception will gradually shift against the seller, eventually reaching the point where, in our experiment, client C3 can avoid the scam, without significant loss of politeness and dignity. Fig. 6.3e and Fig. 6.3f show the evolution of B_{gift} and the seller's public dignity D_p^s for Experiment 3.

6.2.2 $B_{\mathbf{dec}}^{c}$ and $B_{\mathbf{dec}}^{w}$

In Experiment 1, client C1 recognizes the seller's deception after time t=5, which raises B_{dec}^{c} to 0.5. As until time t5 the B_{dec}^{w} value is zero, C1 is not aware of the deception (which will be the ultimate cause of her buying the flower. Clients C2 and C3 recognize the seller's deception through the increase of their respective value of B_{dec}^{w} to 0.3. At time t=12 client C2 already has $B_{\mathrm{dec}}^{w} \approx 0.5$ and $B_{\mathrm{dec}}^{c} \approx 0.5$, which helps him reject those transactions in which the seller was loud and offensive.

Similarly, when the seller approaches client C3, she already knows about the deception with $B_{\mathrm{dec}}^{w} \approx 0.7$, acquired from information from surrounding environment. This helps her reject the offer of the gift and avoid any communication with the seller. However, we can observe that the B_{dec}^{c} of client C3 decreases by 0.05 due to the fact that client had no personal interaction with the seller due to which the decision was solely based upon the information gathered from environment. Fig. 6.4a and 6.4b shows the modeled values of of B_{dec}^{c} of clients and the evolution of B_{dec}^{w} for Experiment 1.
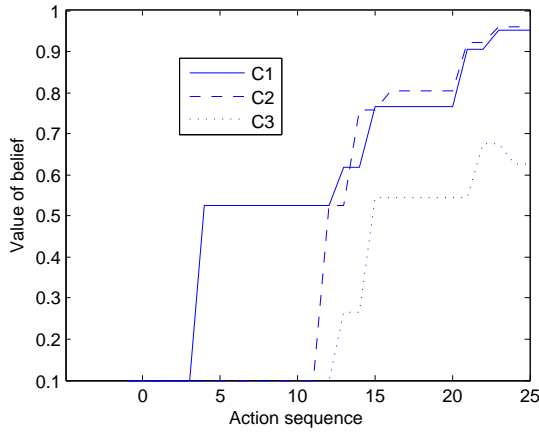
(a) B_{dec}^c of the clients
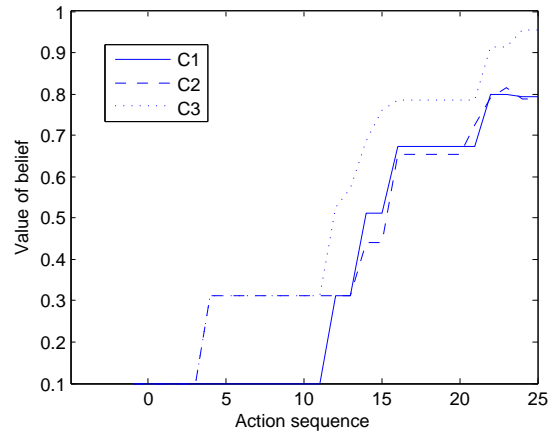
(b) B_{dec}^w of the clients
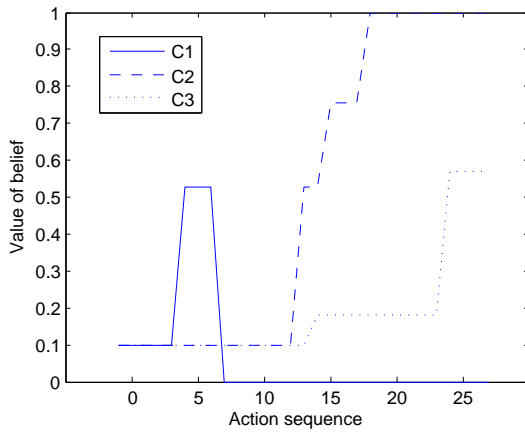
(c) B_{dec}^c of the clients
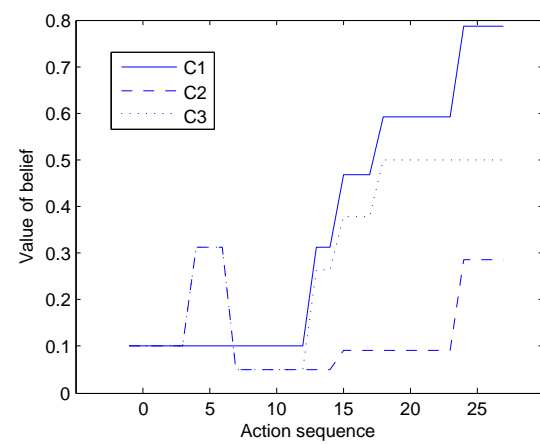
(d) B_{dec}^w of the clients
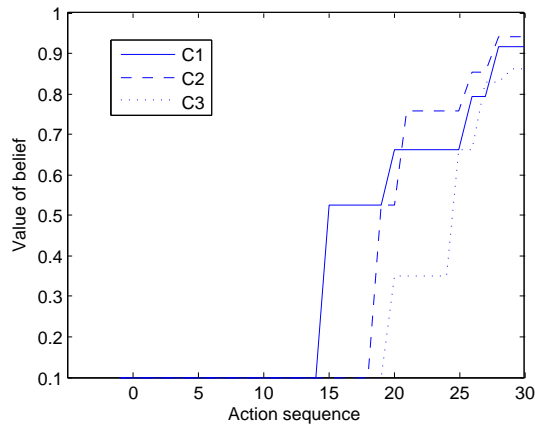
(e) B_{dec}^c of the clients
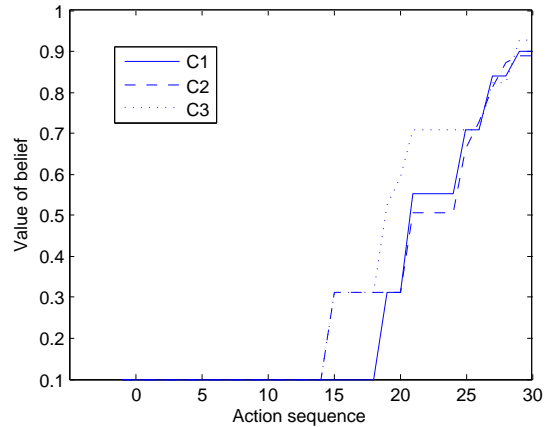
(f) B_{dec}^w of the clients

Figure 6.4: Non-interleaving without breaks (top row), Non-interleaving with breaks (middle row), Interleaving clients (bottom row)

In Experiment 2, the seller waited 20 minutes before approaching the next client. This delay helps the seller to lower the B_{dec}^w. Although the client C2 has high B_{dec}^c as shown in Fig. 6.4c, he does not have sufficient B_{dec}^w (0.3) as shown in Fig. 6.4d to reject the offer publicly. The client C1 has no prior knowledge of seller's deception till time step t3 but after time step t7 this B_{dec}^c is not taken into consideration by other client's B_{dec}^w.

In Experiment 3, C1, C2 and C3 are not aware of the deception, having $B_{\text{dec}}^c = 0$ and $B_{\text{dec}}^w = 0$ until t=15 when the seller is asking C1 for money. Although C1 had witnessed the interaction of the seller with other clients, he had not seen any evidence of deception. Without having the support of the crowd in marking the seller as deceptive, C1 has no argument to reject the payment asked by seller. On the other hand, seeing this, C2 and C3 are rapidly raising their B_{dec}^c and B_{dec}^w values. Client C2 estimates $B_{\text{dec}}^w \approx 0.3$ when asked for the money. However, she judges this as an insufficient support for the crowd to escalate the effort to return the flower. On the other hand, C3 will have a value $B_{\text{dec}}^w \approx 0.7$ when asked for the money at t=23 as shown in Fig. 6.4e. This gives her sufficient confidence on the crowd's support to turn down the seller's offer. Thus, by the end of this interaction, the crowd became aware of the seller's deception. This is also depicted by the loss of the seller dignity D_p^s as shown in the Fig. 6.3f.

CHAPTER 7
SOCIAL CALCULUS - USER STUDY

In previous chapters, we described the foundations of an approach of modeling social-cultural interactions. It is a technique that takes into account the culture-sanctioned social metrics (CSSMs) of the actors as well as the beliefs they hold about specific issues of the current scenario. For CSSMs and Concrete Beliefs, we considered cases when the value is estimated by an actor from the perspective of another actor. We provided a formal definition of the model, described the specific challenges of implementing it in software and described a case study of a complex social interaction. Naturally, this model cannot capture the full richness of human decision making situations - human behavior can be influenced by many conscious decisions, cognitive fallacies, psychological factors and even physiological states. Assigning numbers to social values is an inherently inexact science. However, the working assumption is that the culture enforces a more or less uniform method to calculate the sanctioned social values. This means that we can validate (and, if necessary calibrate) the CSSM model by performing a survey in which persons cognizant with the respective culture will judge the impact on the social values.

A study [57] by Lobato et al. examined human perception of *social signals* based on manipulated sets of *social cues* in a simulated socio-cultural environment. *Social cues* are discrete and observable features of human behavior conveying the social information

embedded in emotional, social and cultural context. A combination of social cues form a *social signal* that describes the perceived meaning of human actions. From the perspective of our model, social signals are cognitive tools that can help in populating CSSMs and interpreting CBs which, in-turn, can be accessed by underlying social cues. A concrete belief of "financial transaction" can be established from social cues such as hand gestures of rubbing fingers with wobbling head movement, whereas angry and threatening social signals will affect the politeness and dignity CSSMs.

In order to provide us with a computational framework, CSSMs must be assigned numerical values. For tangible CSSMs this is an easy task, because they come with their concrete measurement techniques. Thus, worth will be measured in dollars or euros while time will be measured in seconds or minutes.

Things are significantly more complicated for intangible CSSMs. How do we measure politeness or dignity on a numerical scale? While the measurements of such metrics might not necessarily live up to the standards of scientific metrology as practiced in engineering, there is a significant body of work attaching numerical values to intangibles in social settings. The measurement and comparative study of emotional, cultural and social values are regularly done in the social sciences, often using graphical tools such as the interpersonal circumplex to model personality traits [49, 50]. In business and marketing settings it is sometimes important to put a numerical value of the level of politeness of salespeople or customer service [51].

The conclusion is that whether a given metric is a CSSM, whether the intra-cultural uniformity conjecture holds, and how CSSMs are affected by specific interactions can only be validated by asking human users. To perform this, we conducted a user study asking participants to evaluate the CSSM in specific social scenarios. The study was designed to answer the following research questions:

- Which metrics verify the requirements to be a CSSM in a given culture? In order to study this, we asked the users about a relatively wide range of names. These ranged from names which we strongly suspected to be CSSMs (e.g. politeness, compassion) to others that have culturally restricted meaning (e.g. sabr), while others we assumed to be known by a small subset of users (e.g. acedia).

- Verify whether and to what degree the intra-cultural consistency conjecture (Section 3.2.3) holds. This would predict that experimental subjects from the same culture will judge the values of CSSMs similarly.

- Verify whether the CSSM model is relevant in the case of scenarios involving robots. Our assumptions say that the evaluation of the CSSMs are, in principle, learnable. Thus, a robot or software agent can be the evaluator agent in a CSSM. What is not clear, however whether robots can be subject agents of CSSMs when evaluated by humans (can a robot be compassionate?) and whether robots can be perspective agents in the CSSM (would I be compelled to be compassionate if I am in the sight of a robot?). This is achieved by replacing a human with robot in the scenario interaction.

The work in the chapter was performed in collaboration with the Cognitive Sciences Laboratory group at UCF to prepare and perform the study that would contribute to the study of social cues and social signals and provide several new scenarios for the modeling of CSSMs and CBs.

7.1 Method

The story snippets used in the experiments had been designed such that they exhibit cross-cultural interactions, the presence of a robot, as well as emotionally charged social situations. The setting of all the scenarios had been decided to be a Middle-Eastern marketplace in a war-inflicted zone (such as the border between Syria and Turkey). The story actors include a kebab vendor, local market crowd, dejected refugees, and US soldiers manning a security checkpoint. As our survey measures the social and emotional reactions of the survey participants, it was important that the story snippets are emotionally believable. To ensure this, we enrolled the help of an experienced screenwriter, Landon Berry from the Institute of Simulation and Training at UCF. For the survey, we designed two scenarios:

Excited. A buyer/robot (non-refugee) approaches the seller seeking a discount. The buyer/robot is very excited by the arrival of a friend and uses that news to try and secure a discount.

Dejected. The seller is stacking his products at the back of the market when he notices a refugee woman who is trying to convince her traveling companion, a young, male refugee/robot to pass the border with her. The refugee/robot seems unwilling to carry

on and wishes to be left alone. The soldiers manning the checkpoint approach and warn them that the border passing will close soon. The seller approaches the refugee/robot to learn that his friend was killed before they reached market.

In order to study whether the presence of a robot affected the stated CSSM values, for half of the survey participants, the presented scenario involved a young human as the dejected refugee, while for half of them, a humanoid robot. The screenshot of the animation created for *Dejected Scenario* and *Excited scenario* is shown in Fig. 7.1 and Fig. 7.2, respectively. The story script used for designing the scenario animation for *Dejected Scenario* and *Excited scenario* are provided in Appendix A.1 and A.2, respectively.

In the *Excited scenario*, we asked the participants to rate the degree of the CSSMs that are well defined for the Western cultures. These CSSMs include *politeness, dignity, adroitness, compassion.* The *Excited Scenario* CSSMs questions are shown in Table 7.1. In *Dejected Scenario*, we focused on a wider variety of not very obvious CSSMs metrics from different cultures. These are included *Leadership* in US culture, *Acedia* in Catholic culture, *Sabr* in Islamic culture, *Stickler for Rules* in US culture, *Compassion* in US culture, and *Stiff Upper Lip* in British culture. In this scenario, the participant first reads the definition of CSSMs unknown to his culture and then provides feedback on the CSSM questions. Table 7.2 shows the CSSMS questionnaire for the *Dejected Scenario.*

Figure 7.1: Dejected human scenario(top), Dejected robot scenario (bottom).
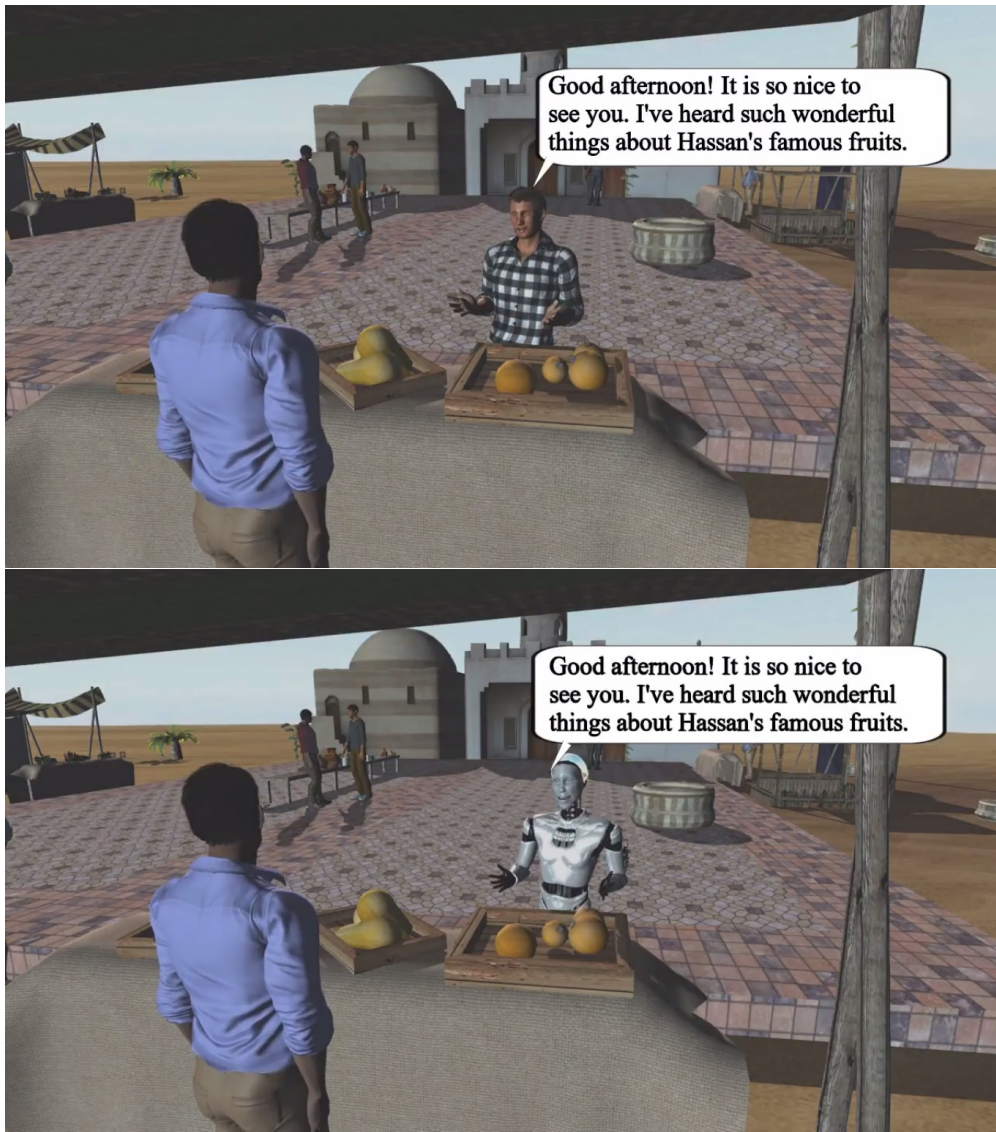
Figure 7.2: Excited human scenario (top), Excited robot scenario (bottom).

The participants for the survey were recruited from UCF graduate students and participants contacted through social networks, and were of a variety of different social and cultural backgrounds. The responses to the survey were collected through the Qualtrics survey management tool. The survey participants received a one-time request by email that

included a link to the survey hosted on a Qualtrics installation at UCF. The identity of the participants was kept anonymous and no incentives were offered for participation in the survey. We collected demographic characteristics including age, gender, race/ethnicity, most spoken language, education status, current profession and geographical location. The survey required approximately 20 minutes of attention from the participants. Each experiment proceeded through the following steps:

- The participant reads a preparatory text about the scenario setting and context.

- The participant watches a short video clip of approximately 2 minutes.

- The participant is debriefed about scenario and asked to evaluate the relevant CSSMs and their effect on the subject actor's actions.

Table 7.1: Survey Questions - Excited Scenario.

Politeness	
Q1	From the Market Vendor's perspective, please indicate your opinion about the degree of politeness that he thinks the crowd believes he is exhibiting.
	CSSM(Generic, Politeness, Vendor, Crowd, Vendor)
Dignity	
Q2	From the Market Vendor's perspective, please indicate your opinion about the degree of dignity that he thinks the crowd believes he is exhibiting.
	CSSM(Generic, Dignity, Vendor, Crowd, Vendor)
Q3	Please indicate your opinion about the degree of dignity that the Customer thinks the crowd believes is exhibited when the customer interacts with the market vendor.
	CSSM(Generic, Dignity, Customer, Crowd, Customer)
Adroitness	
Q4	From the Market Vendor's perspective, please indicate your opinion about the degree to which the Market Vendor thinks the crowd believes him to be an adroit trader.
	CSSM(Generic, Adroitness, Vendor, Crowd, Vendor)

	Continuation of Table 7.1
Q5	From the Market Vendor's perspective, please indicate your opinion about the degree to which the Market Vendor thinks the Customer believes him to be an adroit trader.
	CSSM(Generic, Adroitness, Vendor, Customer, Vendor)
Q6	From the Market Vendor's perspective, please indicate your opinion about the degree to which the Market Vendor believes himself to be an adroit trader.
	CSSM(Generic, Adroitness, Vendor, Vendor, Vendor)
Q7	Please use the scale to indicate how well you understand the term "adroit".
Compassion	
Q8	From the Market Vendor's perspective, please indicate your opinion about the degree of compassion that he thinks the crowd believes he is exhibiting.
	CSSM(Generic, Compassion, Vendor, Crowd, Vendor)
Q9	Please indicate your opinion about the degree of compassion that the Customer thinks the crowd believes is exhibited when the customer interacts with the market vendor.
	CSSM(Generic, Compassion, Vendor, Crowd, Customer)
Q10	From the Market Vendor's perspective, please indicate your opinion about the degree to which the Market Vendor believes himself to be compassionate.

	Continuation of Table 7.1	
	CSSM(Generic, Compassion, Vendor, Vendor, Vendor)	
Q11	Please use the scale to indicate how well you understand the term "compassion".	
Importance of social terms		
Q12	Please use the scale to indicate how important are these terms.	

...

Table 7.2: Survey Questions - Dejected Scenario.

Leadership	
Definition: In US culture, "leadership" is defined as the ability of a person to have a social influence and to be able to enlist the aid and support of others in the accomplishment of a common task.	

Q1	Please indicate your opinion about the degree of leadership refugee Woman exhibits.
	CSSM(US, Leadership, Woman, User, User)
Q2	Please indicate your opinion about what you think refugee Woman believes about her own leadership.
	CSSM(US, Leadership, Woman, Woman, User)
Q3	Please indicate your opinion about what you think refugee Man believes about refugee Woman's leadership.
	CSSM(US, Leadership, Woman, Man, User)

Acedia	
Definition: In Catholic culture, the term "acedia" describes a state of torpor, of not caring or not being concerned with one's position or condition in the world. It can lead to a state of being unable to perform one's duties in life. It is this slothful inability to make decisions.	

	Continuation of Table 7.2
Q4	Please indicate your opinion about the degree of acedia refugee Man exhibits.
	CSSM(Catholic, Acedia, Man, User, User)
Q5	Please indicate your opinion about what you think Refugee Woman believes about refugee Man's acedia.
	CSSM(Catholic, Acedia, Man, Woman, User)
Q6	Please indicate your opinion about what you think the Soldiers believe about refugee Man's acedia.
	CSSM(Catholic, Acedia, Man, Soldier, User)
Sabr	
Definition: The Islamic culture, "sabr" denotes the virtue of being persistent, steadfast and doing good actions even when facing opposition, adversity or calamities.	
Q7	Please indicate your opinion about what you think refugee Woman believes about refugee Man's sabr.
	CSSM(Islamic, Sabr, Man, Woman, User)
Q8	Please indicate your opinion about what you think refugee Man believes about refugee Woman's sabr.
	CSSM(Islamic, Sabr, Woman, Man, User)

Q9	Please indicate your opinion about what you think refugee Woman believes about her own sabr.
	CSSM(Islamic, Sabr, Woman, Woman, User)

Stickler for Rules

Definition: In US slang a "stickler for rules" is a person who believes that rules are very important and they should be followed to the letter all the time.

Q10	Please indicate your opinion about the degree to which you think the Soldiers are sticklers for rules.
	CSSM(US, Stickler for Rules, Soldier, User, User)
Q11	Please indicate your opinion about the degree to which you think refugee Woman believes that the Soldiers are sticklers for rules.
	CSSM(US, Stickler for Rules, Soldier, Woman, User)

Compassion

Definition: In US culture, "compassion" is the response to the suffering of others that motivates a desire to help.

Q12	Please indicate your opinion about the degree of compassion refugee Woman exhibits.
	CSSM(US, Compassion, Woman, User, User)

	Continuation of Table 7.2
Q13	Please indicate your opinion about the degree of compassion the Market Vendor exhibits.
	CSSM(US, Compassion, Vendor, User, User)
Q14	Please indicate the degree to which you think refugee Woman believes the Market Vendor is compassionate.
	CSSM(US, Compassion, Vendor, Woman, User)
Q15	Please indicate the degree to which you think the Market Vendor believes refugee Woman is compassionate.
	CSSM(US, Compassion, Woman, Vendor, User)

Stiff Upper Lip

Definition: In British culture, the term "stiff upper lip" refers to one who displays fortitude in the face of adversity, or exercises great self-restraint in the expression of emotion.

Q16	Please indicate the degree to which you think refugee Woman believes refugee Man has a stiff upper lip.
	CSSM(British, Stiff Upper Lip, Man, Woman, User)
Q17	Please indicate your opinion about what you think refugee Woman believes about her own stiff upper lip.
	CSSM(British, Stiff Upper Lip, Woman, Woman, User)

Continuation of Table 7.2	
Q18	Please indicate the degree to which you think refugee Man believes refugee Woman has a stiff upper lip.
	CSSM(British, Stiff Upper Lip, Woman, Man, User)

7.2 Discussion

7.2.1 Survey Subject Statistics

We gathered 97 responses representing various countries and cultures. Fig. 7.3(b), shows the distribution of the participants by their location at the time of taking the survey. Fig. 7.3(a), shows the distribution of the participants by the language they speak most of the time. The respondents from India provided various language names which are combined as Indian language in the plot. Fig. 7.3(c) shows the participant's contribution to the survey by gender. USA and India provided the highest number of respondents for comparing the impact of cultural differences on the CSSMs. Most of the survey respondents have education level of bachelor's or higher with either primary or secondary language as English.
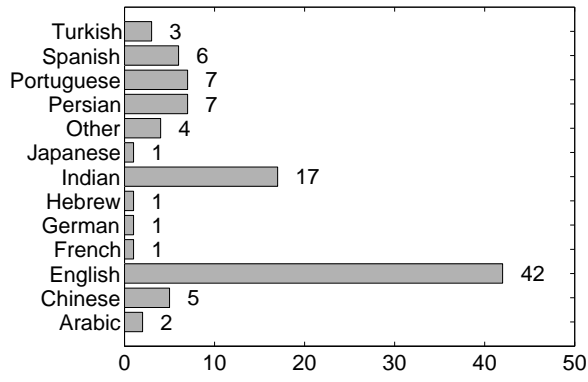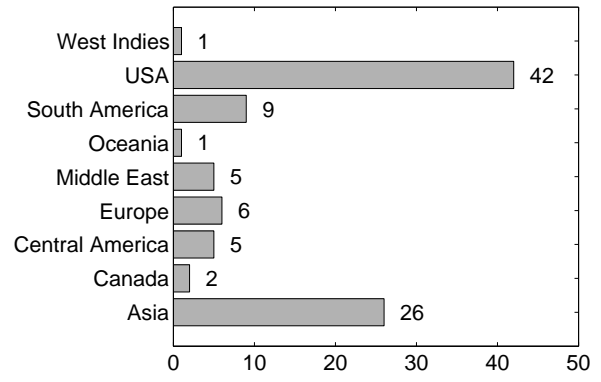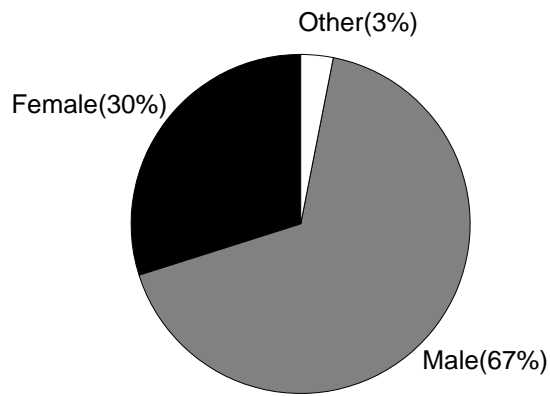
Figure 7.3: Survey participants geographical distribution. (a) By Language (b) By Location (c) By Gender

7.2.2 CSSMs relevant to the scenarios

The characters in the *Dejected Scenario* a variety of emotions involving sadness, urgency, humanity, duty and necessity. The survey used the following list of candidate terms for CSSMs:

- **Leadership:** In US culture, "leadership" is defined as the ability of a person to have a social influence and to be able to enlist the aid and support of others in the accomplishment of a common task.

- **Acedia:** In Catholic culture, the term "acedia" describes a state of torpor, of not caring or not being concerned with one's position or condition in the world. It can lead to a state of being unable to perform one's duties in life. It is a slothful inability to make decisions.

- **Sabr:** The Islamic culture, "sabr" denotes the virtue of being persistent, steadfast and doing good actions even when facing opposition, adversity or calamities.

- **Stickler for Rules:** In US slang a "stickler for rules" is a person who believes that rules are very important and they should be followed to the letter all the time.

- **Compassion:** In US culture, "compassion" is the response to the suffering of others that motivates a desire to help.

- **Stiff Upper Lip:** In British culture, the term "stiff upper lip" refers to one who displays fortitude in the face of adversity, or exercises great self-restraint in the expression of emotion.

(a)   (b)
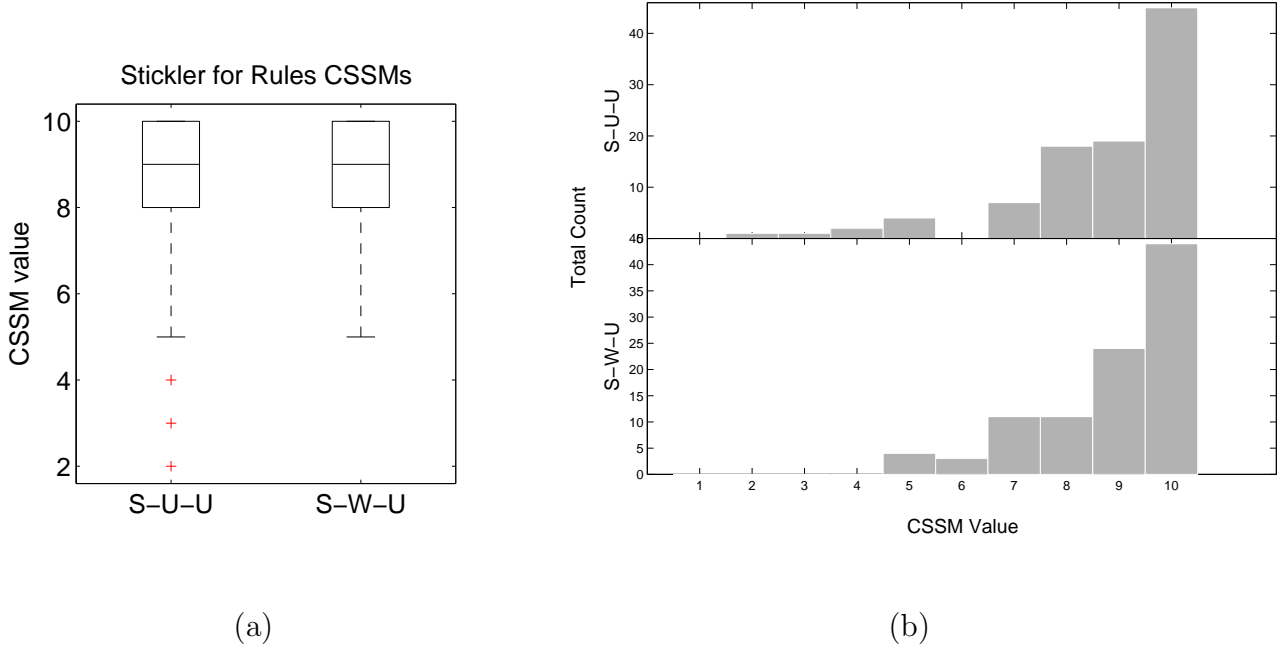
Figure 7.4: Response to 'Stickler for Rules' CSSM where S-U-U = CSSM(US, Stickler for Rules, Soldier, User, User), S-W-U = CSSM(US, Stickler for Rules, Soldier, Woman, User)

One question to investigate is what CSSMs are appropriate for this scenario modeling? For instance, a certain measurable property denoted by a name might not be a CSSM at all, in the sense that the culture does not provide sufficient rules to evaluate it. Second, even if a metric is a CSSM, certain combinations of the subject, perspective, and estimator might not be feasible and/or useful in a given scenario. For instance, the subject might not perform actions that allow us to evaluate the given CSSM, or the perspective actor might not see

those actions. On the other hand, we found that certain metrics, turned out to provide very consistent evaluations and are likely to be also usable in other scenarios. As shown in Fig. 7.4, the 'Stickler for Rules' metric, although derived from a US slang expression, do not create any confusion in the mind of respondents. A very large fraction of respondents selected a very high value for both CSSMs. Thus, although the expression was initially unknown to many respondents, they had no difficulty in evaluating it after learning it through definition.

(a)   (b)
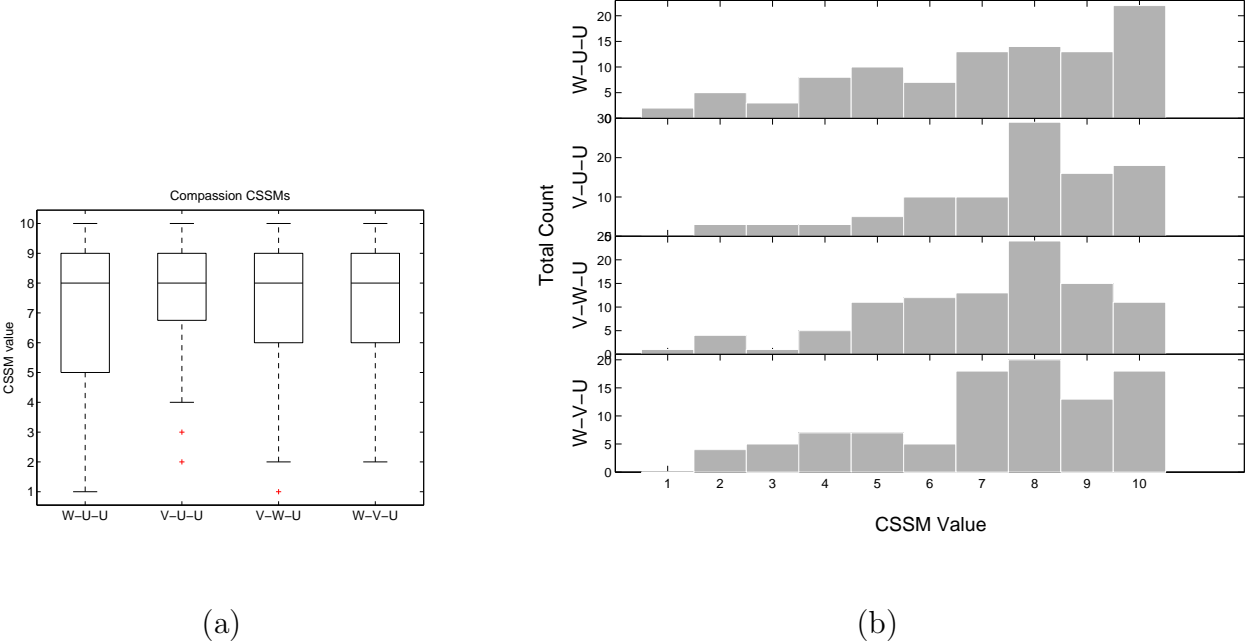
Figure 7.5: Response to 'Compassion' CSSM where W-U-U = CSSM(US, Compassion, Woman, User, User), V-U-U = CSSM(US, Compassion, Vendor, User, User), V-W-U = CSSM(US, Compassion, Vendor, Woman, User), W-V-U = CSSM(US, Compassion, Woman, Vendor, User)

While 'Stickler for Rules' provided a more consistent evaluation than we expected, we obtained the opposite result for the metric of 'Compassion'. Compassion is a widely used term, extensively studied in psychology, often seen as a religious commandment, and extolled as part of school curriculum. When the participants were asked whether they understand the term, they have consistently responded with very high confidence. On the other hand, the evaluation results shown in Fig. 7.5, show a very wide spread of values, illustrating the ambiguity in the evaluation. Clearly, without further qualifications, 'Compassion' would be a poor choice of metric for the CSSM framework. A possible approach would be to break the term 'Compassion' into more precise forms, such as 'Exhibit kindness toward the sufferer' and 'Provide help to the sufferer'. For instance, in our scenario, the vendor didn't offer any help or support during the interaction, but respondents provided higher degree of 'Compassion' to him than to the female companion. An interesting aspect of the responses was the female respondents had evaluated the participants to have a higher level of compassion than male responders Fig. 7.6.
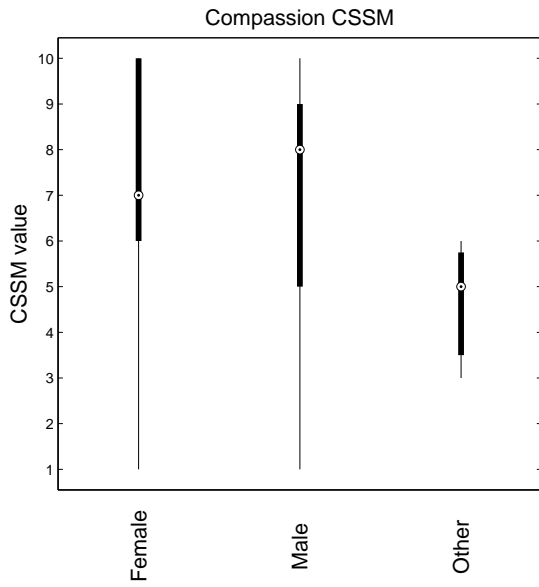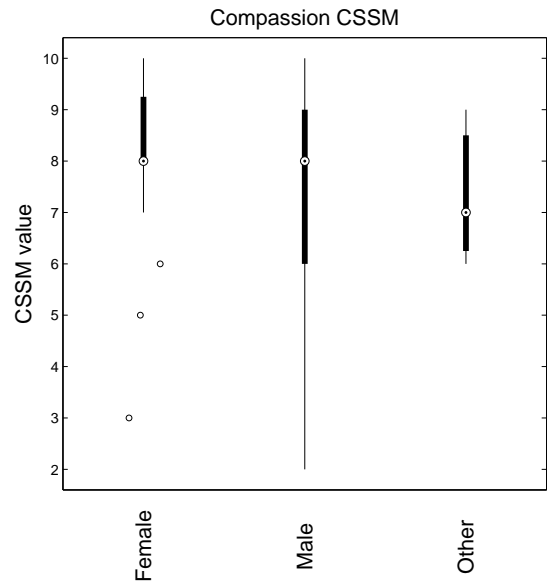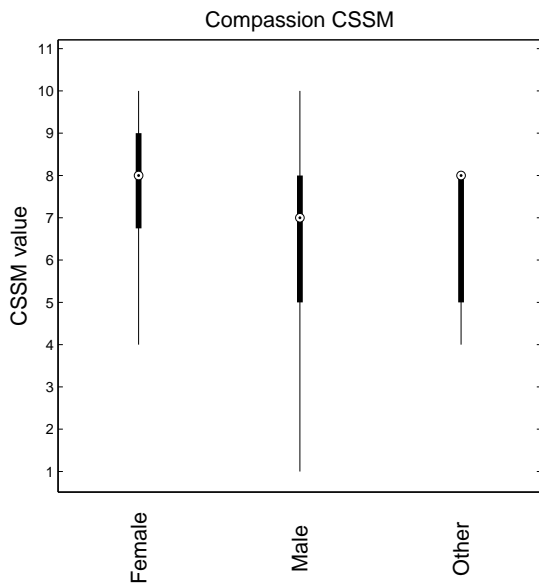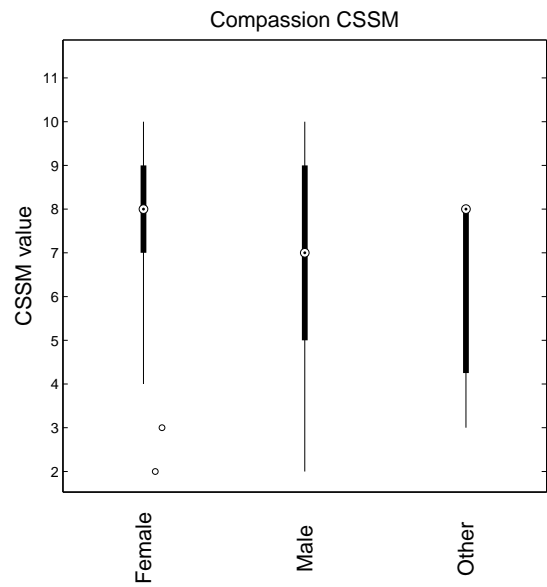
Figure 7.6: Comparison of responses to 'Compassion' CSSM by gender where (a) CSSM(US, Compassion, Woman, User, User), (b) CSSM(US, Compassion, Vendor, User, User), (c) CSSM(US, Compassion, Vendor, Woman, User), (d) CSSM(US, Compassion, Woman, Vendor, User)

7.2.3   Comparison of Human vs Robot

(a)   (b)   (c)
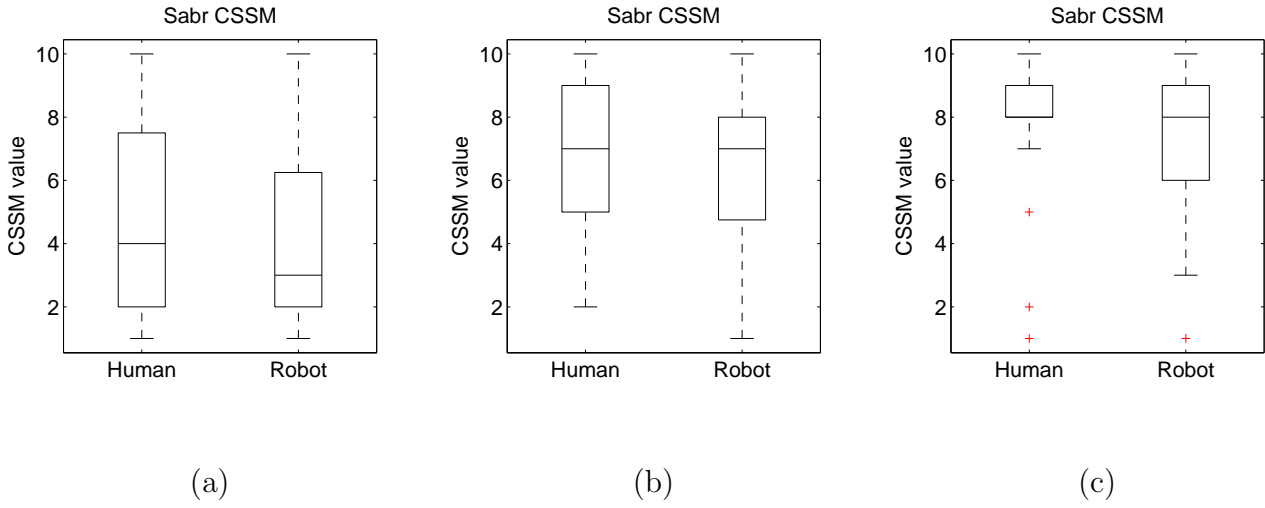
Figure 7.7: Response to 'Sabr' CSSM in Human vs Robot scenario, where (a) CSSM(Islamic,

Sabr, Man/Robot, Woman, User), (b) CSSM(Islamic, Sabr, Woman, Man/Robot, User), (c)

= CSSM(Islamic, Sabr, Woman, Woman, User)
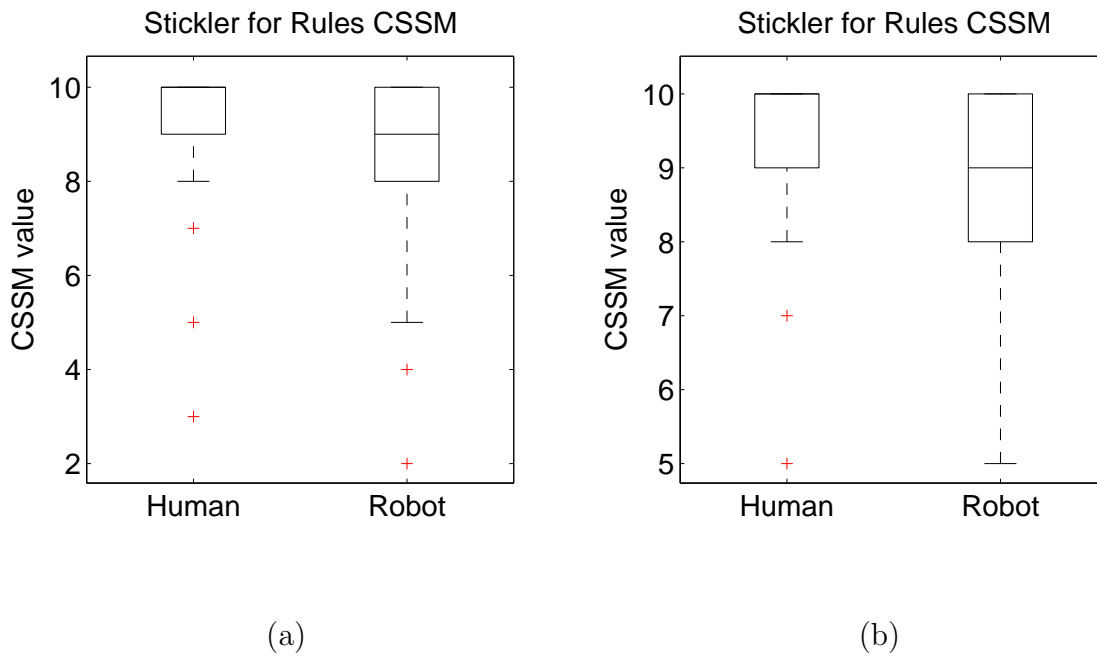
Figure 7.8: Response to 'Stickler for Rules' CSSM in Human vs Robot scenario, where (a) CSSM(US, Stickler for Rules, Soldier, User, User), (b) CSSM(US, Stickler for Rules, Soldier, Woman, User)

A robot or software agent can be the evaluator agent in a CSSM. What is not clear, however whether robots can be subject agents of CSSMs when evaluated by humans (can a robot be compassionate?) and whether robots can be perspective agents in the CSSM (would I be compelled to be compassionate if I am in the sight of a robot?). In order to get responses to these statements, an equal number of the users were shown videos with a humanoid robot replacing the human actor in the Dejected and Excited scenarios.

We found that the simple presence of the robot in the scenario has created confusion in people's judgment about the metric. For an instance, Fig. 7.9(c) shows the comparison of

responses to the 'Sabr' metric in human vs robot scenario where respondents were asked to evaluate the 'Sabr' of Woman from her own perspective. We found that the CSSM values have a wider spread when the robot was present in the scene, although the CSSM does not refer to the robot. A similar effect was found in the case of the 'Sticker for rules' CSSMs as shown in the Fig. 7.8.

Interesting results had been obtained when comparing values for the compassion of the woman and vendor through various perspectives Fig. 7.9. Although none of these values use the robot as a subject, perspective and evaluation agent, due to the logic of the story, the compassion of the woman and vendor is expressed with respect to the dejected person or robot. It turns out that in this case the answers are more tightly grouped in the robot compared to the human case. This confirms our hypothesis that CSSMs are applicable to model scenarios involving robots. For completeness sake, we need to mention that the videos used a humanoid robot that could replicate the gestures of the human participant. The results might be different for a visibly non-humanoid robot.

Another aspect of the results is that the study participants had consistently evaluated the compassion of the woman and seller higher when the dejected person was a human than when it was a robot. This result is surprising, because it implies that the participants had held the woman and seller to higher compassion standards when evaluated with respect to a robot, compared to a human.
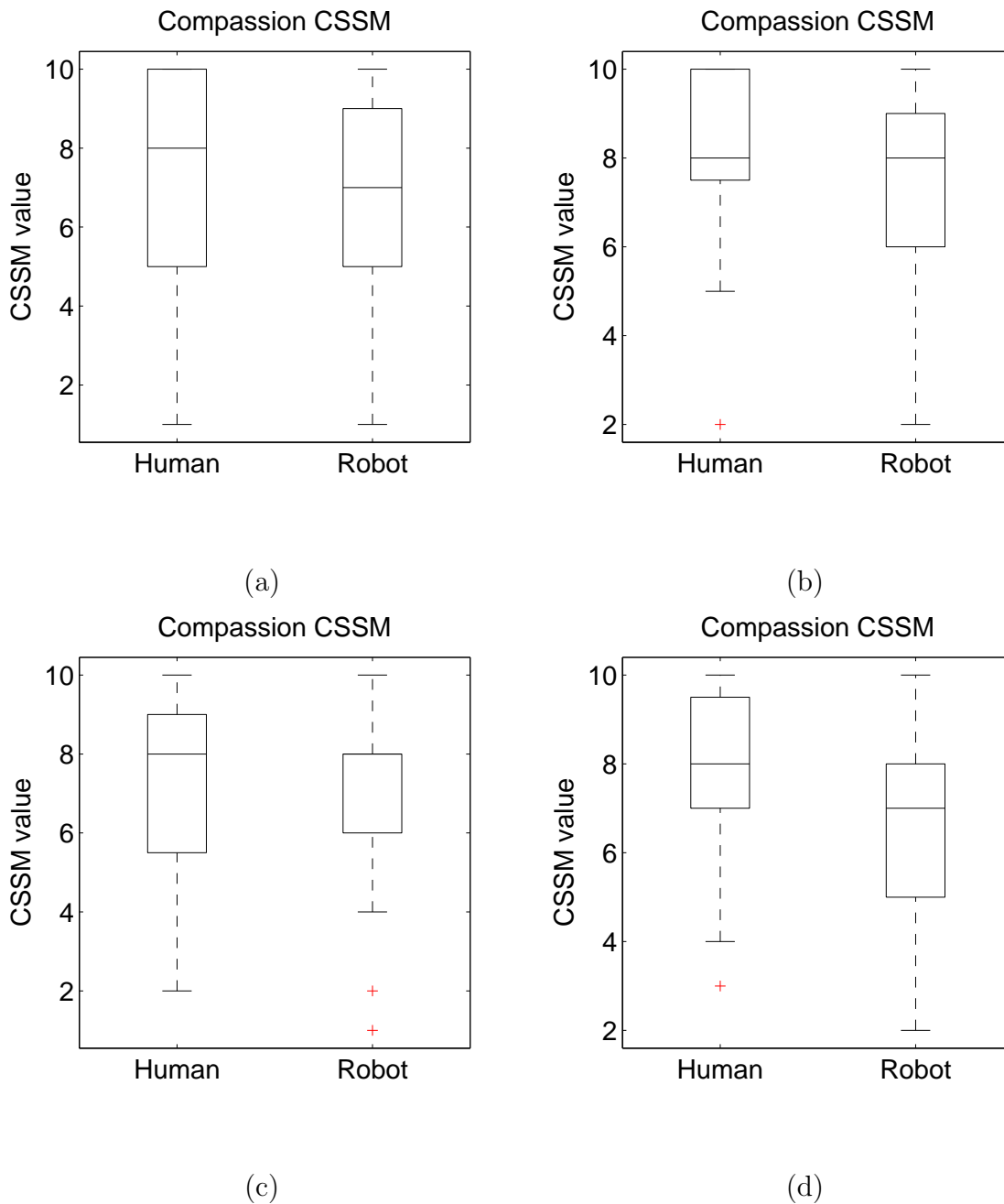
Figure 7.9: Response to 'Compassion' CSSM in Human vs Robot scenario, where (a) CSSM(US, Compassion, Woman, User, User), (b) CSSM(US, Compassion, Vendor, User, User), (c) CSSM(US, Compassion, Vendor, Woman, User), (d) CSSM(US, Compassion, Woman, Vendor, User)

7.2.4 Evaluation of CSSMs accross multiple cultures

The multiplication of possible perspectives increases the complexity of the CSSM evaluation. If we need to consider different models of evaluating the CSSMs for every social agent, the framework would have no practical utility.

The *intra-cultural uniformity conjecture* states that we don't need to consider different evaluation models on the individual basis: it is enough to model them once for every culture. Two persons who have the same information and use the same algorithm for the evaluation provided by the shared culture should reach approximately the same value for a CSSM. We need to emphasize that this does not mean that different individuals in the same culture will *behave* the same way. One social agent might follow the rules of politeness while another might not - the conjecture only says that they would both be *aware* of the rules.

In the following, we will study how the culture affected the responses of the participants in the survey. The first challenge is that it is difficult to determine what culture(s) the participants are aware of. As we discussed previously in this dissertation, the same word might represent different CSSMs in different cultures. All the respondents to the survey were English-speakers (although not necessarily as a first language). As it was not practical to ask the participants in what culture they are providing their answers, we had used their stated first language and geographical location as a proxy of their culture.

The survey participants were distributed unevenly among the various language groups and geographical locations. The two largest sets of responders in the survey were English lan-

guage speakers from USA and various regional language speakers from Indian subcontinent. The remaining groups were in general too small to generate meaningful average results.

Fig. 7.10 shows the response to the 'Sabr' CSSM of the woman seen from her own and the dejected man's perspective, broken down based on the first language of the responders. We find that in this case, the responders from the two largest language provided remarkably consistent responses.
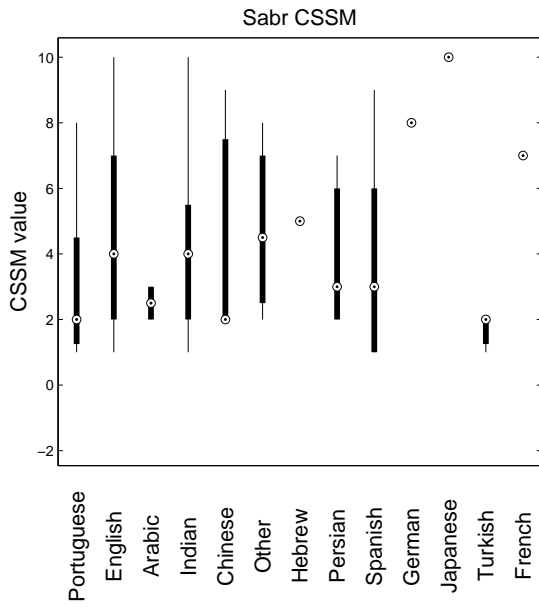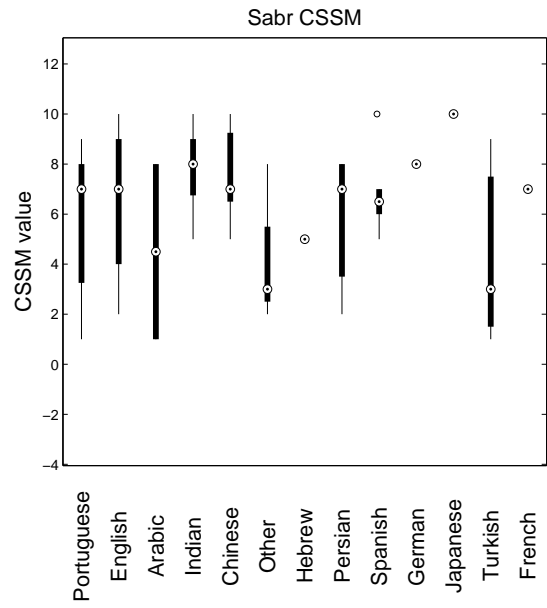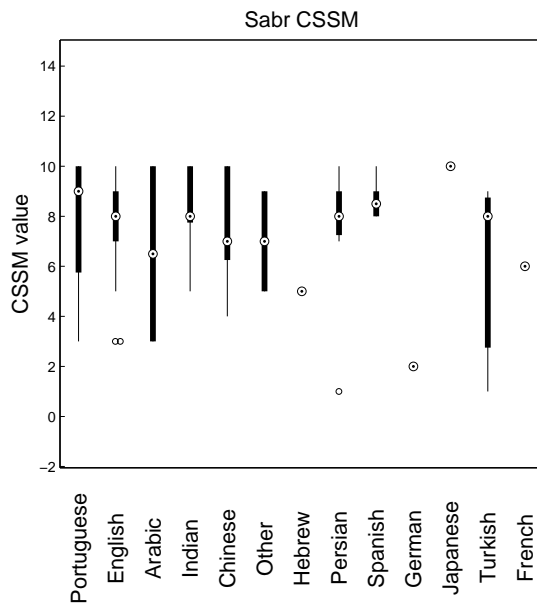
Figure 7.10: Response to 'Sabr' CSSM where (a) CSSM(Islamic, Sabr, Man, Woman, User),
(b) CSSM(Islamic, Sabr, Woman, Man, User), (c) CSSM(Islamic, Sabr, Woman, Woman,
User)

7.2.5   Are people aware of the CSSM?

In contrast to many other factors of social life that humans consider without conscious consideration, CSSMs are hypothesized to be explicit, conscious and even requiring significant cognitive effort. If these hypotheses are true, this means that the humans users should be able to introspect on these features.

To verify the participant's self-perception about CSSMs, as the last question of the survey, we asked users to provide the degree of importance they give to metrics such as politeness, dignity, and compassion to assess the situation and consider when deciding their next action. The results are shown in Fig. 7.11. Overall, the responders had answered with very high values to this question.

(a)   (b)
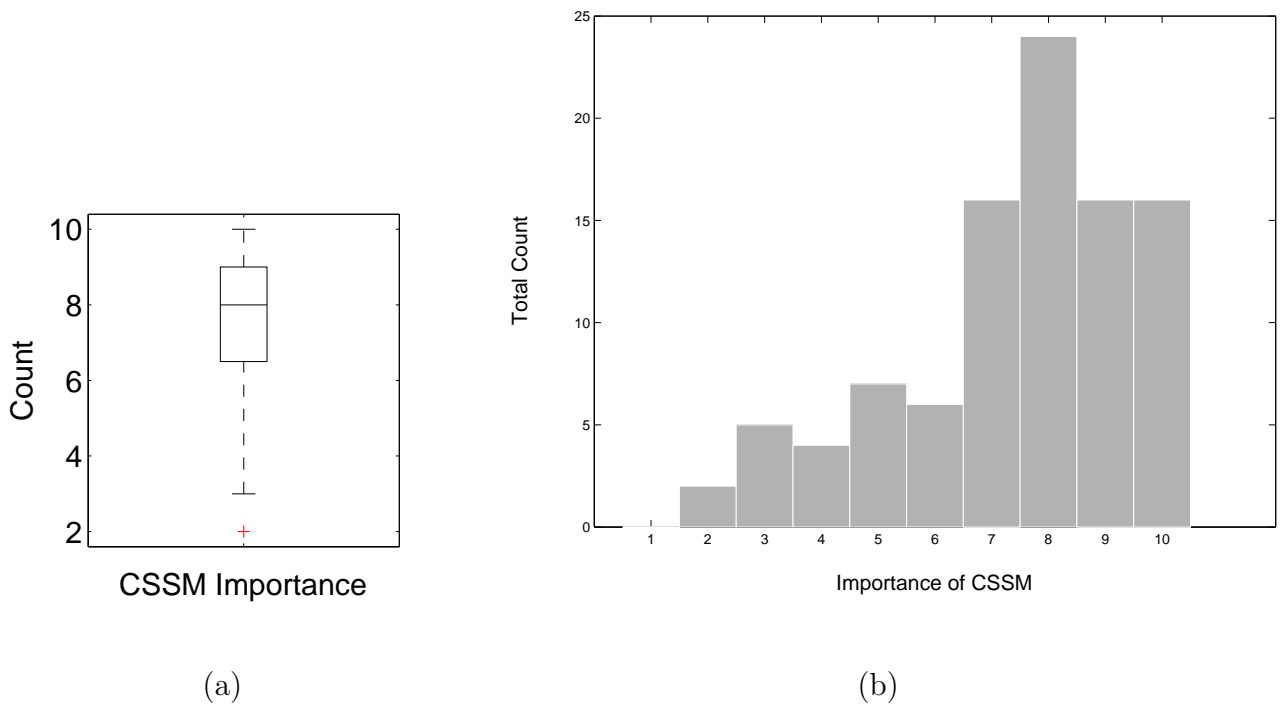
Figure 7.11: Participants response to Q12: Importance of the social terms in the scenario.

7.2.6   Representativeness of the survey

One of the important considerations is the representativeness of the survey: are the results of the survey representative of the CSSMs of the target population? It is well known that many academic surveys suffer from the problem of using respondents who are in many ways divergent from the general population and are, in certain ways, "weird" [58]. In the following we will discuss some of the obstacles we perceive in the representativeness of our results.

- The culture of the survey takers might not be an exact match of the target culture. This is an unavoidable bias - for a perfect localization, one would need to use respondents from the exact geographical location we model.

- The distorting factor of social class: the survey subjects have been drawn from a significantly higher social strata (students, engineers, doctors) than the average composition of the market. It is to be determined whether the social class affects the calculations of CSSMs. Our conjecture is that it has only a minimal effect, through secondary implications, which we will outline below.

- The impact of persons cognizant of multiple cultures. Many of the respondents have received some level of Western or Western-style education. It is to be determined whether this impacts their evaluation of the CSSMs. Our conjecture is that is at most a minimal impact. We assumed that people cognizant of multiple cultures are able to evaluate separate CSSMs according to multiple cultures (naturally, within the limit of the cognitive load they can handle). Then, they decide which CSSM-dependent rules of conduct apply in the current situation (which might be a combination of rules), and plan their actions in function of (not necessarily in obeisance to) these rules. This behavior model implies that even people who do not follow rules according to these CSSM settings, will still be able to calculate them.

CHAPTER 8
CONCLUSION AND FUTURE WORK

This thesis describes the foundation and applications of an approach which promises to model social-cultural situations using a technique which takes into account the culture-sanctioned social metrics of the actors as well as the belief that they hold about specific issues of the current scenario. For both types, we are considering cases when the value is estimated by an actor from the perspective of another actor. We provided a formal definition of the model, described the specific challenges of implementing it in software, described case studies of a complex social interaction and its implementation on multiple scenarios. Naturally, this model can not capture the full richness of human decision making situations - human behavior can be influenced by many conscious decisions, cognitive fallacies, psychological factors and even physiological states. We argue, however, that our model represents a step forward from simplistic models of one-dimensional utility maximization. For instance, in our case study of the Spanish Steps scenario we successfully explained behaviors which, from a utility maximization perspective would appear irrational.

For our future work, we plan further refinements of the model, and extensive applications for many practical scenarios. A significant challenge is to find ways to reduce the need for manual knowledge engineering in the development of the models. One natural research direction is to develop a library of behaviors for specific cultures and situations, which will

allow us to assemble models for new scenarios faster. Finally, we plan to deploy our model

in specific applications such as conversational user interfaces and mobile robots [59].

APPENDIX A
SCENARIO SCRIPTS

A.1   Scenario Script - Dejected

EXT. Kobane Market - Late Afternoon

We are in Kobane, a Syrian border town just outside of Turkey. We see a small, ad hoc Turkish-style market, pieced together from scrap wood and rubble. Tapestries and (in most cases) strips of muslin are strung to metal poles, signifying individual booths. The structure isn't ideal, but it's functional given the circumstances. Merchants are packing up their wares for the night and loading them onto old trucks and trailers. The market is becoming increasingly deserted.

Behind the market looms an expansive, military checkpoint. Jutting out from the concrete structure are rows of barbed wire fences, stretching off into the distance. Central to the structure is a towering, metal gate, above which stand U.S. and Turkish troops, armed with rifles and non-lethal sensory lasers.

Jersey barriers topped with barbed wire create two distinct paths leading up to the gate. 200 meters from the gate are bright yellow, metal speed bumps, underneath which are housed metal "fangs" capable of sending 150,000 volts to the undercarriage of a vehicle, rendering it useless. The contrast in security, stability, and power between the market and military structure is staggering.

At the corner of the market, closest to the military structure, we see a lone kebab vendor, Hassan, putting away the last of his produce. As he walks behind his station, he sees a lone refugee, AI, slumped against a wall. A female refugee, Amena, approaches him.

The story snippets used in the experiments had been designed such that they exhibit cross-cultural interactions, the presence of a robot, as well as emotionally charged social situations. The setting of all the scenarios had been decided to be a Middle-Eastern marketplace in a war-inflicted zone (such as the border between Syria and Turkey). The participants include a kebab vendor, local market participants, dejected refugees, and US soldiers manning a security checkpoint. As our survey measures the social and emotional reactions of the survey participants, it is important that the story snippets are emotionally believable. To ensure this, we enrolled the help of an experienced screenwriter, Landon Berry. We are currently working on creating four different scenarios:

AMENA. I just spoke to one of the soldiers. They're closing the entry down for the night, but if we hurry, we can make it through.

AI. (no response)

AMENA. We don't have much time. I need for you to come with me. Please.

AI. You go.

AMENA. I don't want to leave without you. Let's go.

AI. (no response)

AMENA. I know how hard this is, but you have to think of yourself now, okay?

Hassan approaches.

HASSAN. Is everything okay?

AMENA. We need to cross the border, but he won't budge.

AI. I'm not leaving him behind.

AMENA. You have no choice.

AI. I won't leave him.

AMENA. He's dead, don't you understand that?! There's nothing for you here except for a slow death as you freeze during the night.

AI. Just leave me. Some things are worse than dying.

AMENA. (under her breath) Tozz feek.

HASSAN. I'm sorry for your loss. Can I ask who died?

AMENA. His friend, Sayid.

HASSAN. How did it happen?

AMENA. A car bomb went off when we were passing Al Hasakah.

> **We see a soldier round a corner and make his way over to Amena and AI.**

SOLDIER. If you're still planning on going, you need to go now. We won't be able to wait any longer.

AMENA. But none of that matters now, because we need to cross the border!

AI. I don't expect you to understand.

SOLDIER. You've got five minutes until we close down. If you can't make it, then you'll want to look for a place to stay for the night. Curfew starts after nightfall. You'll need to get off the streets before then.

We see the soldier turn and march away.

AMENA. Wait! Please!

HASSAN. Do you have a place to stay tonight?

AMENA. No. We arranged for a pickup just across the border, but if we don't go now, we'll miss it.

AI. I just... can't.

AMENA (to Hassan). Please, do something! We don't have much time!

We see night approaching quickly. As AI turns his eyes once again to the ground, Hassan is left with a choice: persuade AI, or hurry to his own abode before curfew.

A.2   Scenario Script - Dejected

EXT. Kobane Market - Afternoon

We are in Kobane, a Syrian border town just outside of Turkey. We see a small, ad hoc Turkish-style market, pieced together from scrap wood and rubble. Merchants are piling bowls of bright orange and red spices onto tables, and vendors are heating various meats and vegetables over fire pits. Tapestries and (in most cases) strips of muslin are strung to metal poles, signifying individual booths. The structure isn't ideal, but it's functional given the circumstances. Behind the market stands a military checkpoint. Two traffic lanes, corralled by Jersey barriers, lead up to a large, yellow traffic arm. American and Turkish soldiers, armed with rifles, patrol the block around the gate. A large, and very well-armored military vehicle stands adjacent to the gate. On this side of the city, there is no other way in or out.

At the corner of the market, closest to the checkpoint, we see a lone kebab vendor, Hassan, stoking the coals of a fire pit. Long skewers of lamb sizzle above the flames. A line of market-goers wait in front of his stand, looking forward to the best kebabs in Kobane.

CUSTOMER 1. Thanks!

HASSAN. Please enjoy.

CUSTOMER 2. Two please.

HASSAN. That will be eight pounds.

CUSTOMER 2. Perfect. Thank you!

We now see a very enthusiastic customer, Tarek/AI, fall in line at Hassan's stand. While the rest of the customers seem to be going about their normal routines, Tarek's lavish hand gestures and constant fidgeting set him apart.

HASSAN. Good afternoon. What can I get you?

CUSTOMER 3. Just one please.

HASSAN. Here you go. Four pounds please.

CUSTOMER 3. Many thanks!

HASSAN. Good afternoon. What can I get you?

TAREK. Good afternoon! It's so nice to see you! I've heard such wonderful things about Hassan's kebabs. I simply had to find out for myself!

HASSAN. That's very kind of you.

TAREK. You see, I'm meeting my dear friend today, and the best way I know to greet him is to arrive with what I understand are the best kebabs in Kobane.

HASSAN. Wonderful! So, two kebabs then?

TAREK. Yes please! The best two kebabs you have!

HASSAN. Here you go. That will be eight pounds.

We see Tarek pull out a few pounds. He counts them and then begins to pat down all of his pockets.

TAREK. I'm so terribly embarrassed, but I seem to only have five pounds.

HASSAN. That's alright. Just one then?

TAREK. But I'm so excited to bring these two wonderful kebabs to my friend. Is there any way I could get a discount?

HASSAN. Unfortunately...

TAREK. What if I tell everyone I see on my way to meet my friend that they need to come and try for themselves Hassan's famous kebabs?!

HASSAN. Well...

TAREK. I would be most honored to spread the name of the best kebab stand in Kobane! What do you think? Please?

Hassan is left with a choice: give Tarek the discount, or ask that he pay in full for a single kebab.

APPENDIX B
SURVEY STATISTICS

B.1   Dejected Scenario

B.1.1   Overall response to CSSMs

(a)

(b)
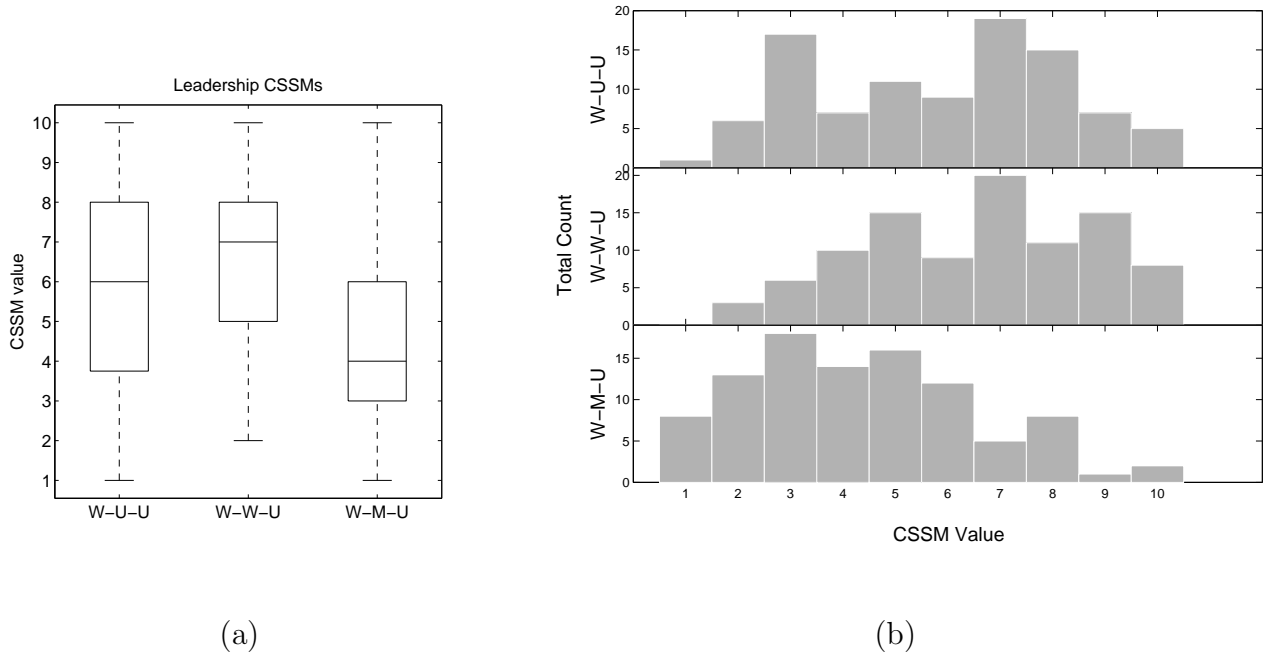
Figure B.1: Response to 'Leadership' CSSM where W-U-U = CSSM(US, Leadership, Woman, User, User), W-W-U = CSSM(US, Leadership, Woman, Woman, User), W-M-U = CSSM(US, Leadership, Woman, Man, User)
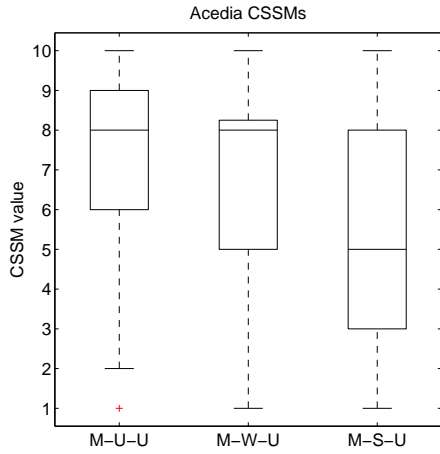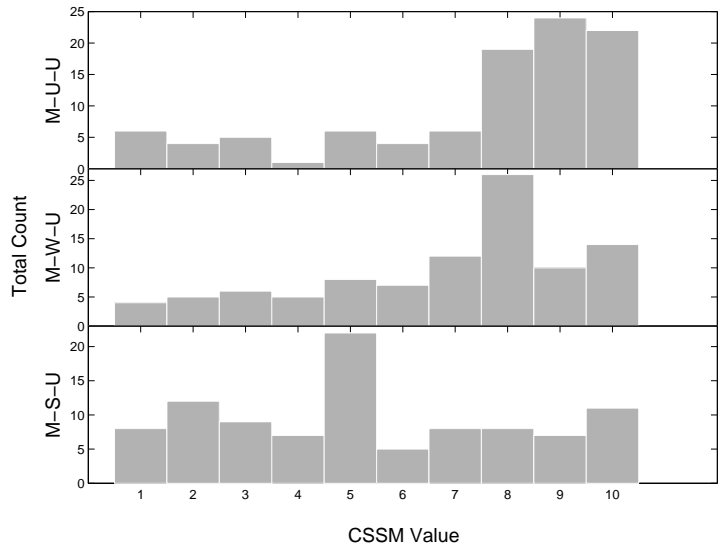
|	
(a)	(b)

Figure B.2: Response to 'Acedia' CSSM where M-U-U = CSSM(Catholic, Acedia, Man, User, User), M-W-U = CSSM(Catholic, Acedia, Man, Woman, User), M-S-U = CSSM(Catholic, Acedia, Man, Soldier, User)
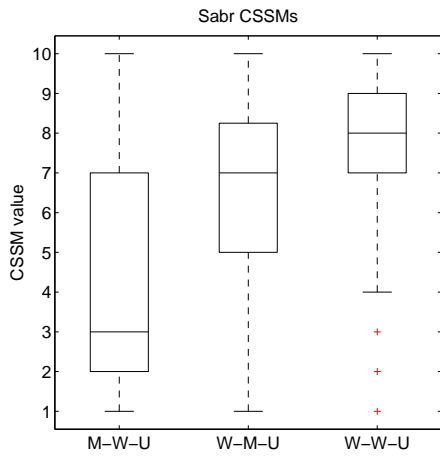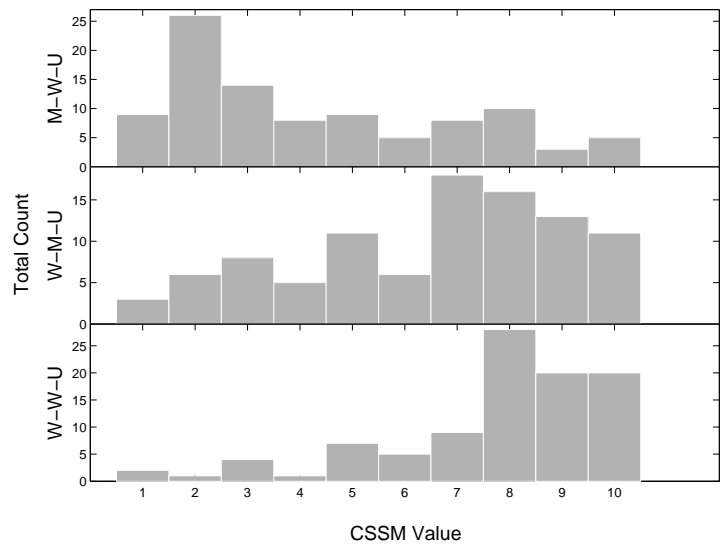
(a)   (b)

Figure B.3: Response to 'Sabr' CSSM where M-W-U = CSSM(Islamic, Sabr, Man, Woman, User), W-M-U = CSSM(Islamic, Sabr, Woman, Man, User), W-W-U = CSSM(Islamic, Sabr, Woman, Woman, User)

162
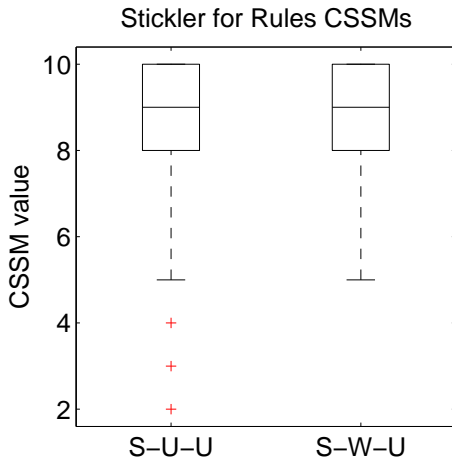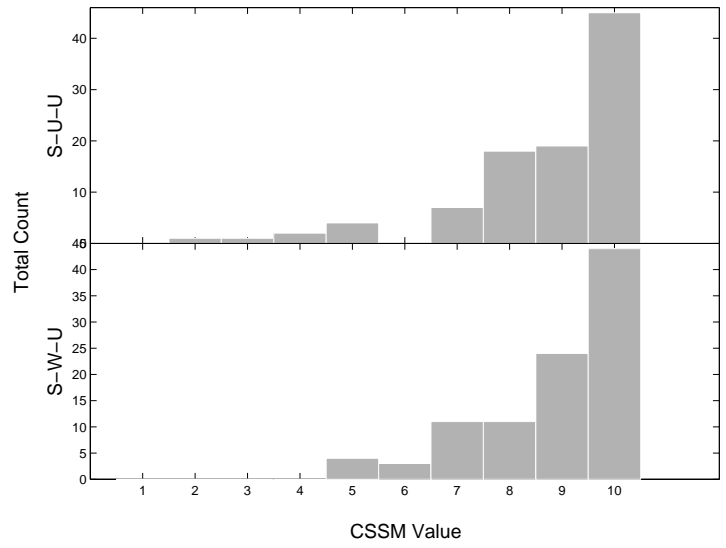
(a)   (b)

Figure B.4: Response to 'Stickler for Rules' CSSM where S-U-U = CSSM(US, Stickler for Rules, Soldier, User, User), S-W-U = CSSM(US, Stickler for Rules, Soldier, Woman, User)
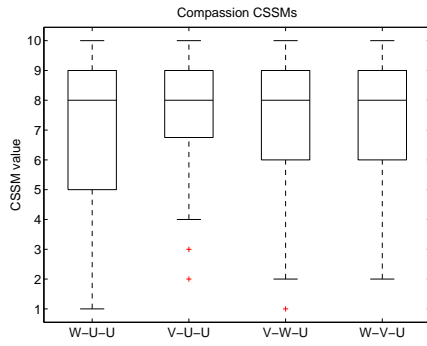
(a)
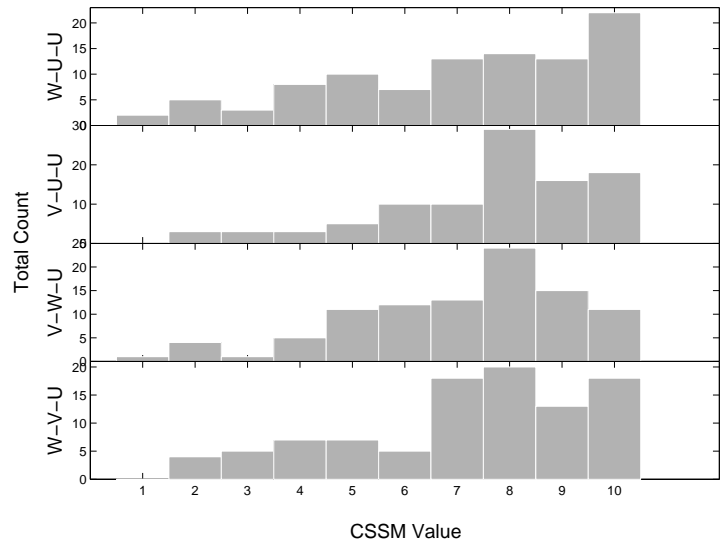
(b)

Figure B.5: Response to 'Compassion' CSSM where W-U-U = CSSM(US, Compassion, Woman, User, User), V-U-U = CSSM(US, Compassion, Vendor, User, User), V-W-U = CSSM(US, Compassion, Vendor, Woman, User), W-V-U = CSSM(US, Compassion, Woman, Vendor, User)
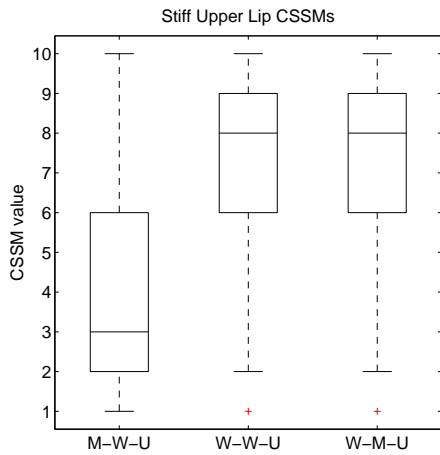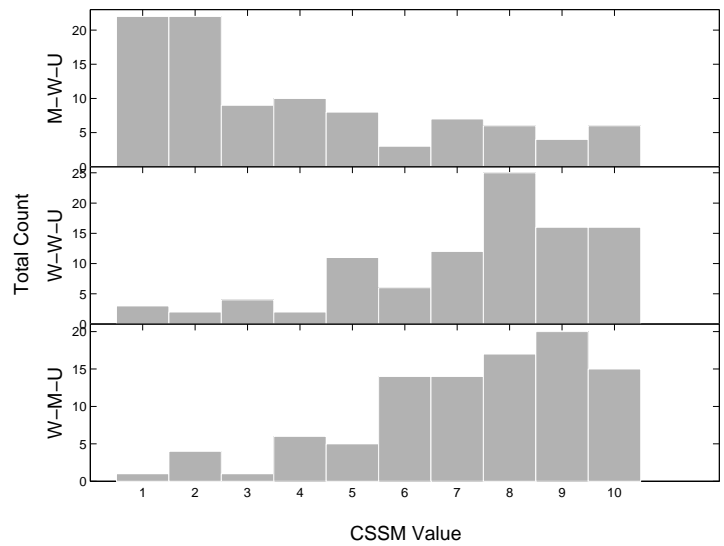
(a) (b)

Figure B.6: Response to 'Stiff Upper Lip' CSSM where M-W-U = CSSM(British, Stiff Upper

Lip, Man, Woman, User), W-W-U = CSSM(British, Stiff Upper Lip, Woman, Woman, User),

W-M-U = CSSM(British, Stiff Upper Lip, Woman, Man, User)

B.1.2 Comparison of Human to Robot scenario

(a)   (b)   (c)
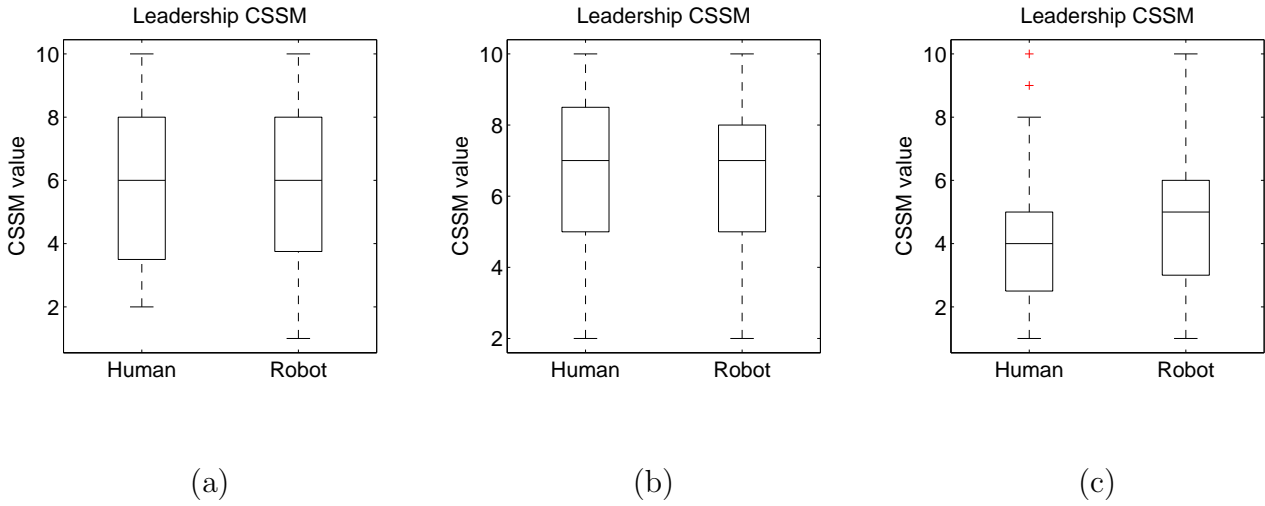
Figure B.7: Response to 'Leadership' CSSM in Human vs Robot scenario, where (a) CSSM(US, Leadership, Woman, User, User), (b) CSSM(US, Leadership, Woman, Woman, User), (c) CSSM(US, Leadership, Woman, Man/Robot, User)

(a)   (b)   (c)
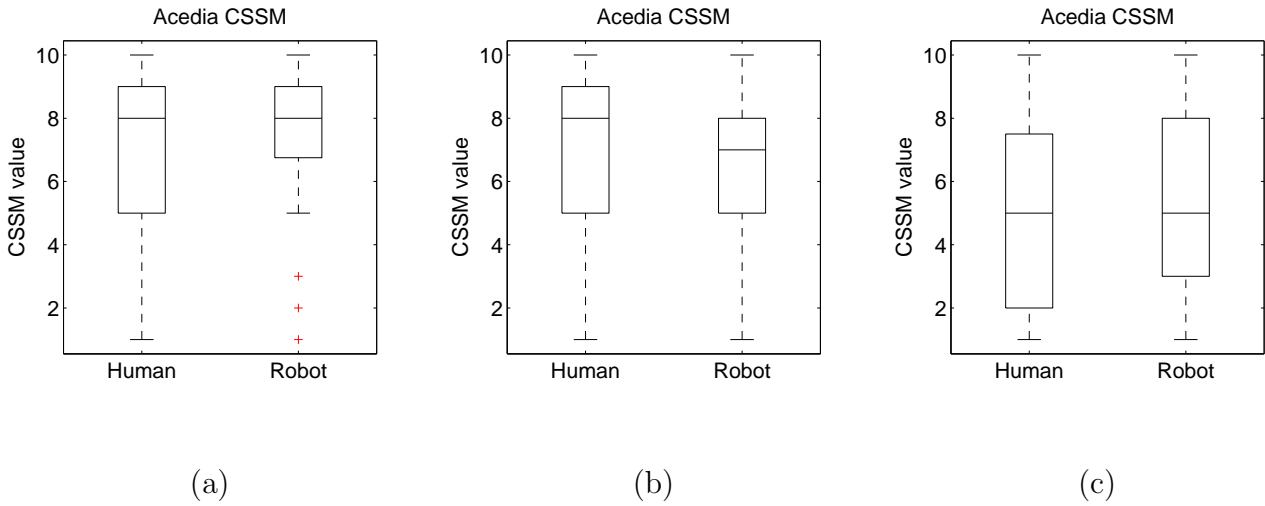
Figure B.8: Response to 'Acedia' CSSM in Human vs Robot scenario, where (a) CSSM(Catholic, Acedia, Man/Robot, User, User), (b) CSSM(Catholic, Acedia, Man/Robot, Woman, User), (c) CSSM(Catholic, Acedia, Man/Robot, Soldier, User)

(a)   (b)   (c)
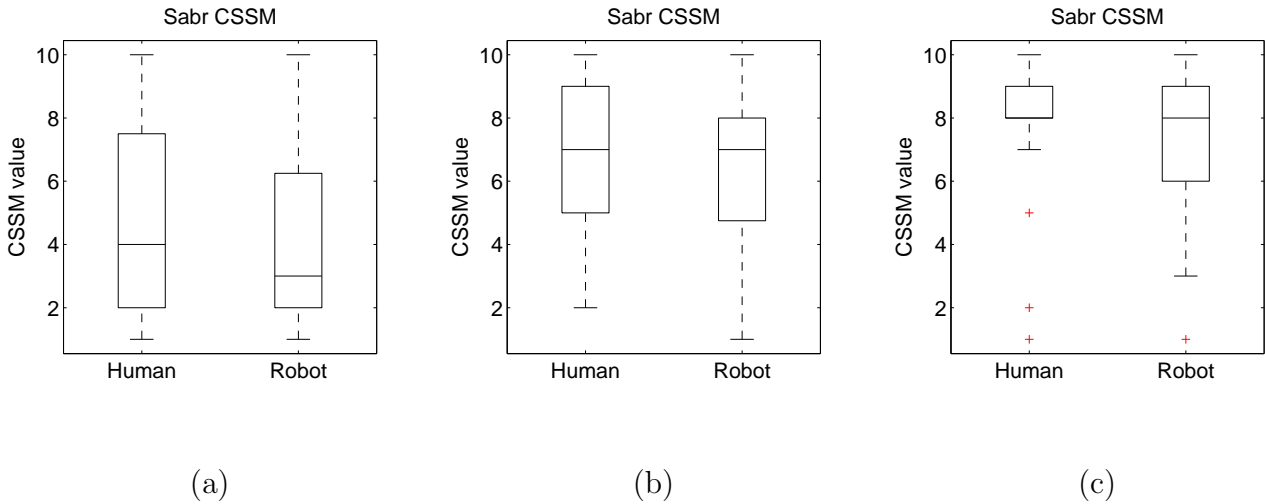
Figure B.9: Response to 'Sabr' CSSM in Human vs Robot scenario, where (a) CSSM(Islamic, Sabr, Man/Robot, Woman, User), (b) CSSM(Islamic, Sabr, Woman, Man/Robot, User), (c) = CSSM(Islamic, Sabr, Woman, Woman, User)

167
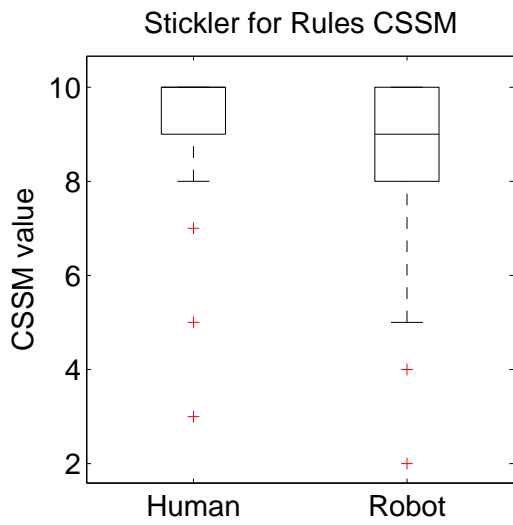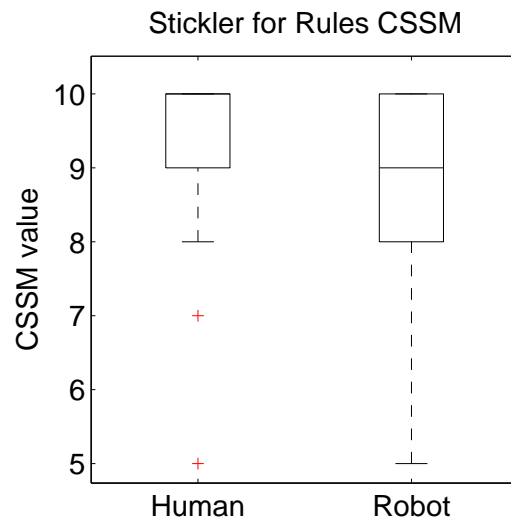
(a)   (b)

Figure B.10: Response to 'Stickler for Rules' CSSM in Human vs Robot scenario, where (a) CSSM(US, Stickler for Rules, Soldier, User, User), (b) CSSM(US, Stickler for Rules, Soldier, Woman, User)
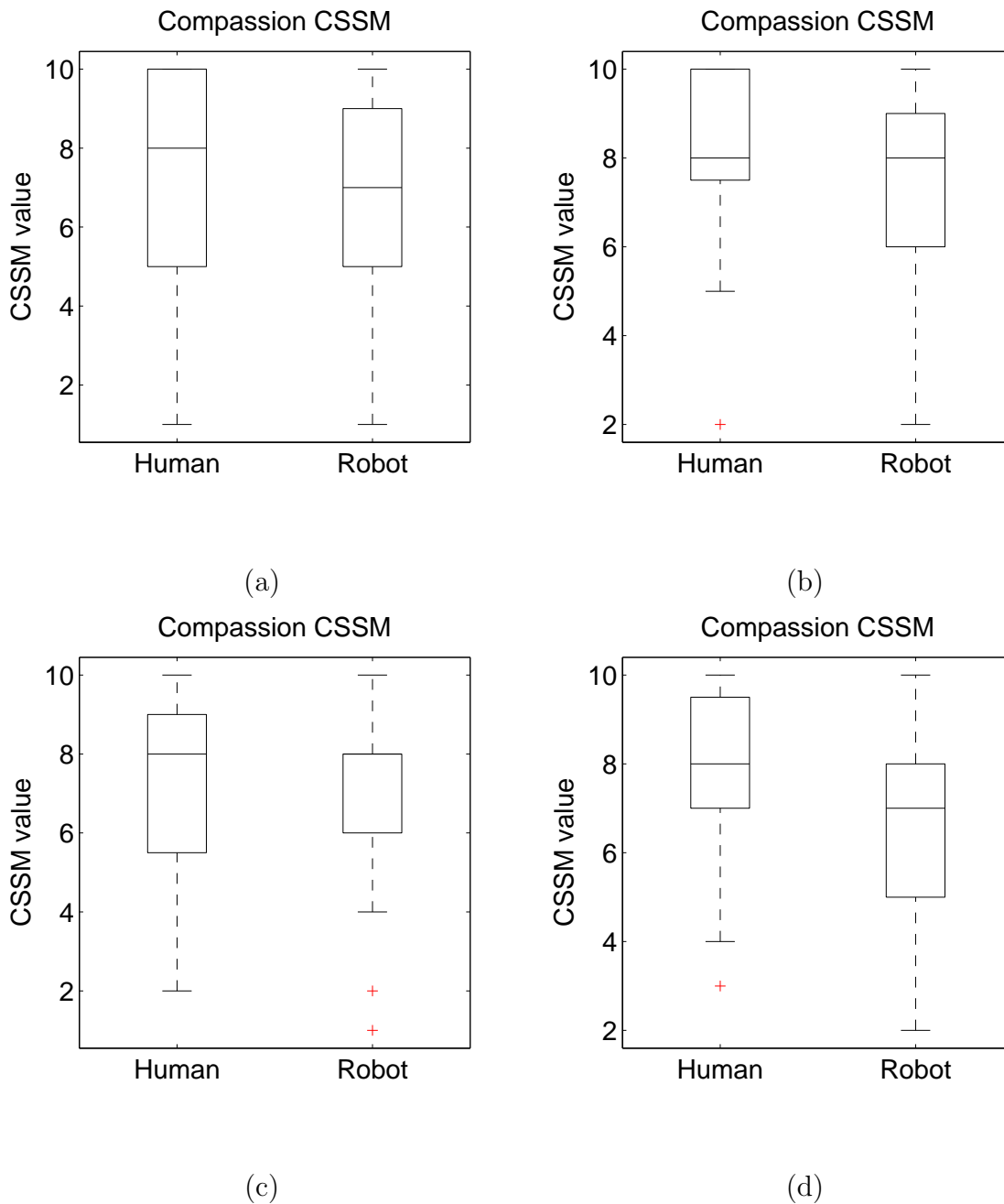
Figure B.11: Response to 'Compassion' CSSM in Human vs Robot scenario, where (a) CSSM(US, Compassion, Woman, User, User), (b) CSSM(US, Compassion, Vendor, User, User), (c) CSSM(US, Compassion, Vendor, Woman, User), (d) CSSM(US, Compassion, Woman, Vendor, User)
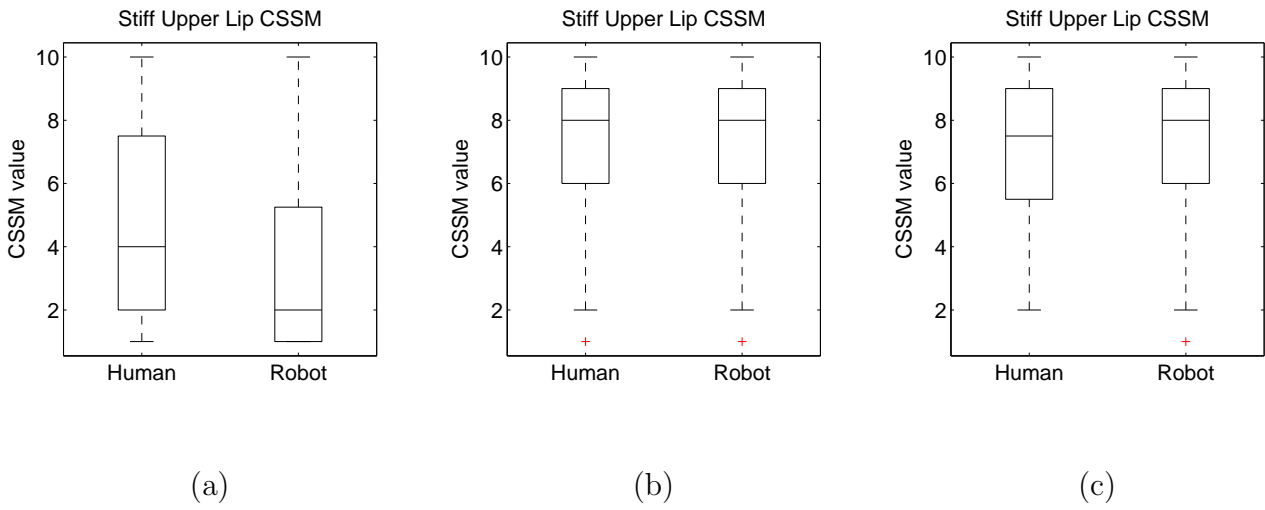
Figure B.12: Response to 'Stiff Upper Lip' CSSM in Human vs Robot scenario, where (a) CSSM(British, Stiff Upper Lip, Man/Robot, Woman, User), (b) CSSM(British, Stiff Upper Lip, Woman, Woman, User), (c) CSSM(British, Stiff Upper Lip, Woman, Man/Robot, User)

B.1.3 Comparison of CSSM evaluation categorized by participant language
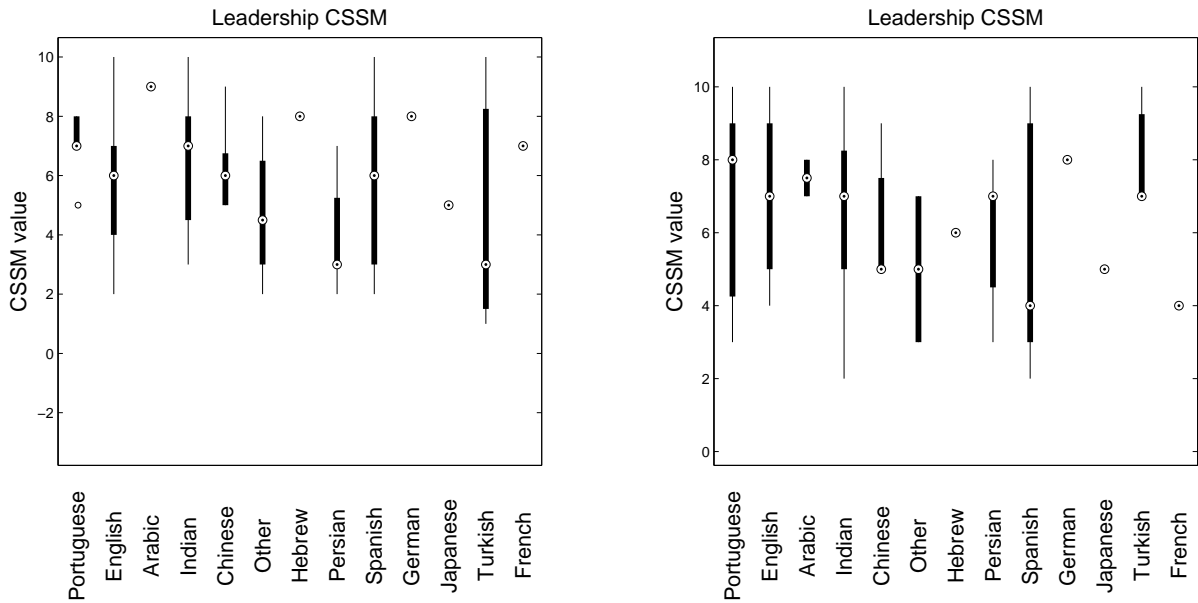
(a)

(b)
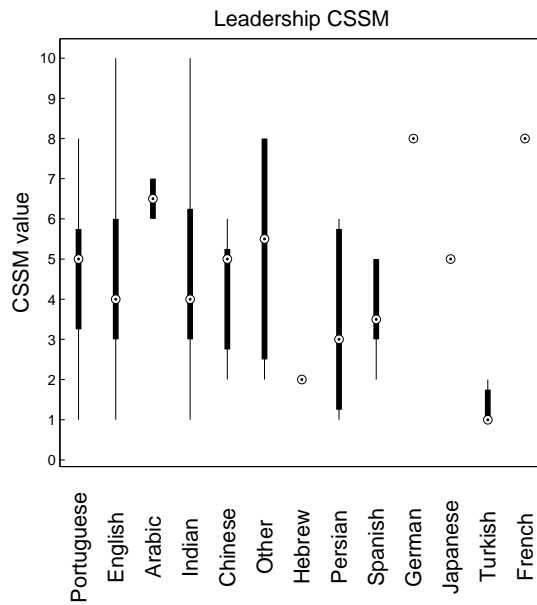
(c)

Figure B.13: Response to 'Leadership' CSSM where (a) CSSM(US, Leadership, Woman, User, User), (b) CSSM(US, Leadership, Woman, Woman, User), (c) CSSM(US, Leadership, Woman, Man, User)
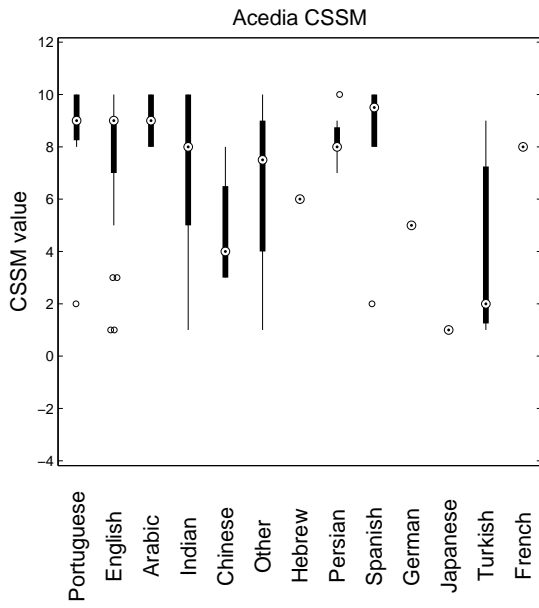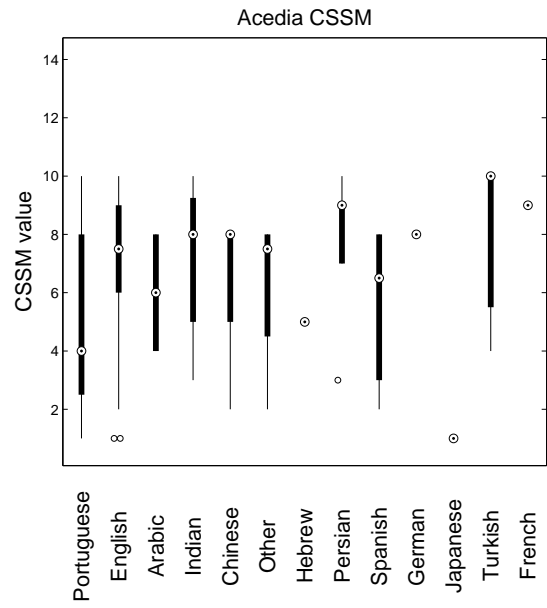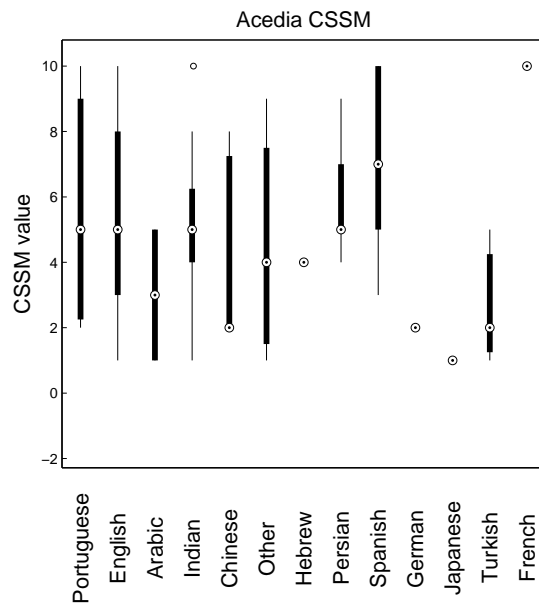
Figure B.14: Response to 'Acedia' CSSM where (a) CSSM(Catholic, Acedia, Man, User, User), (b) CSSM(Catholic, Acedia, Man, Woman, User), (c) CSSM(Catholic, Acedia, Man, Soldier, User)
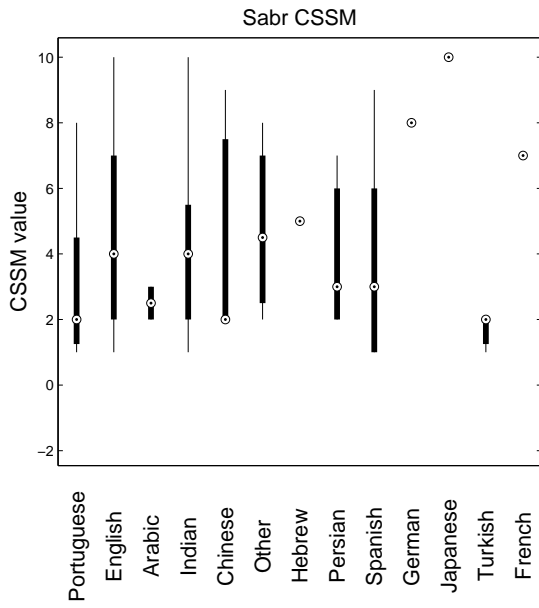
(a)
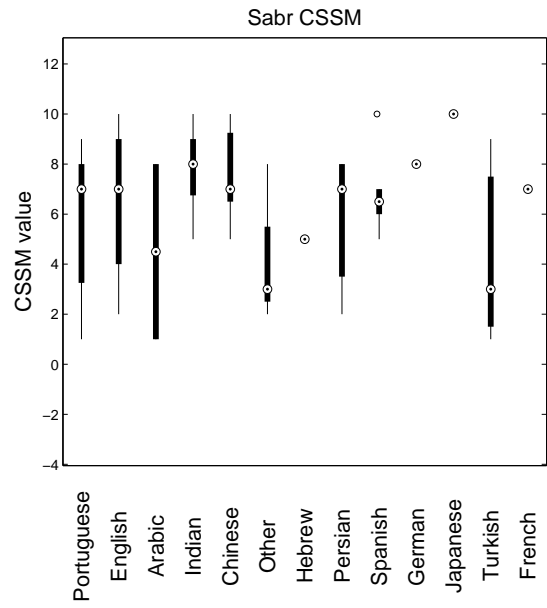
(b)
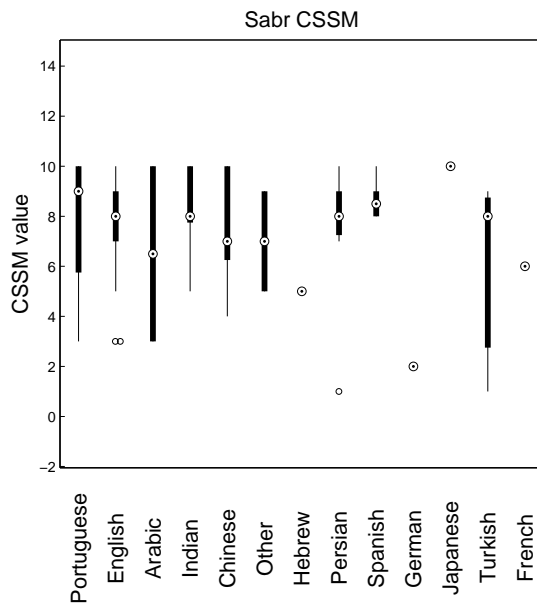
(c)

Figure B.15: Response to 'Sabr' CSSM where (a) CSSM(Islamic, Sabr, Man, Woman, User),
(b) CSSM(Islamic, Sabr, Woman, Man, User), (c) CSSM(Islamic, Sabr, Woman, Woman,
User)

174
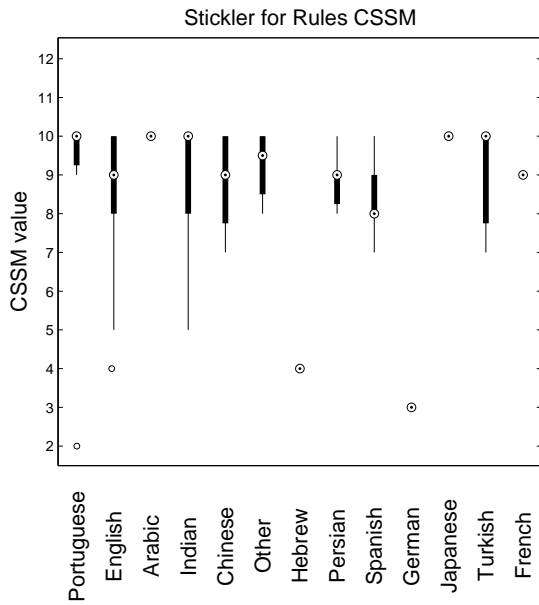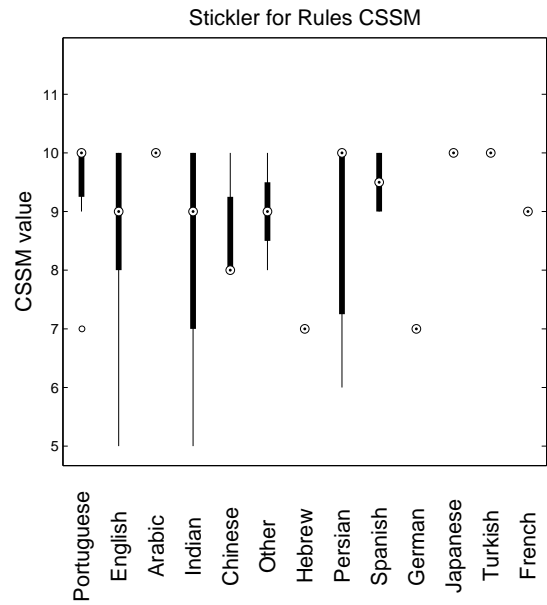
Figure B.16: Response to 'Stickler for Rules' CSSM where (a) CSSM(US, Stickler for Rules, Soldier, User, User), (b) CSSM(US, Stickler for Rules, Soldier, Woman, User)
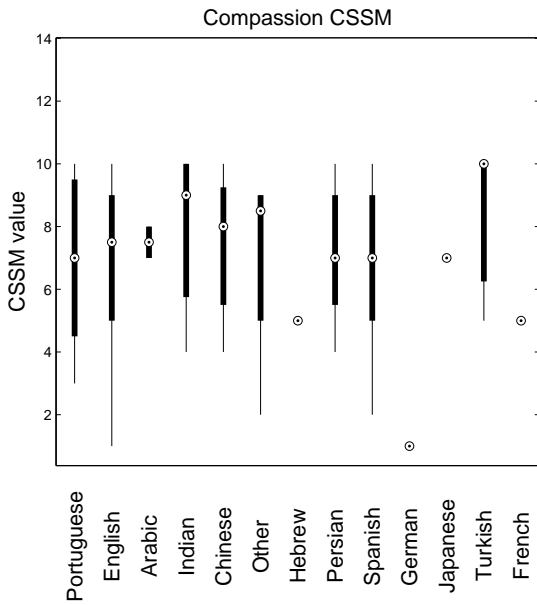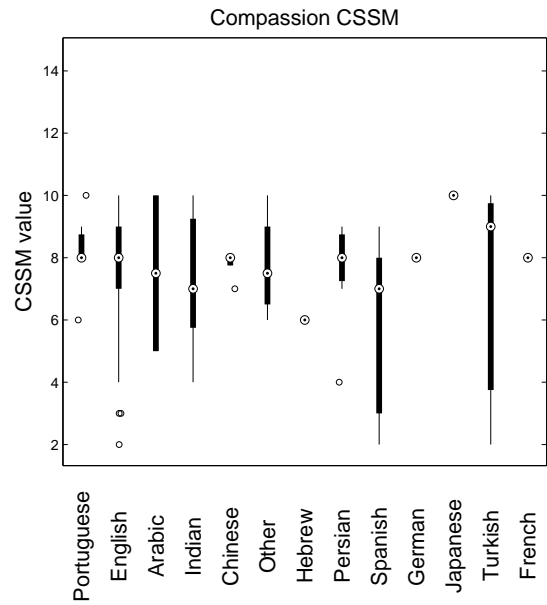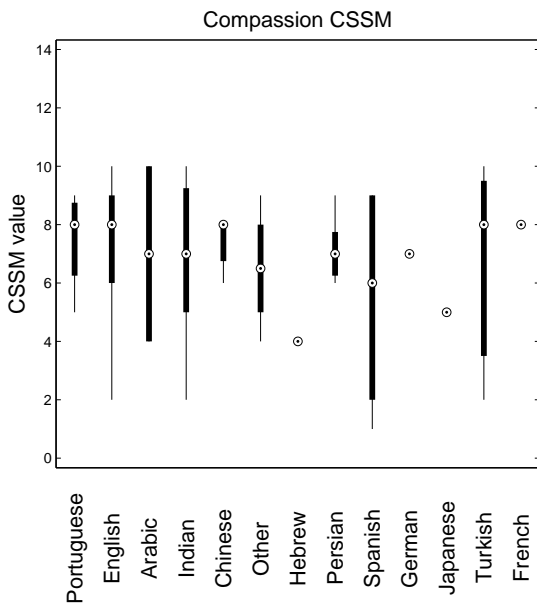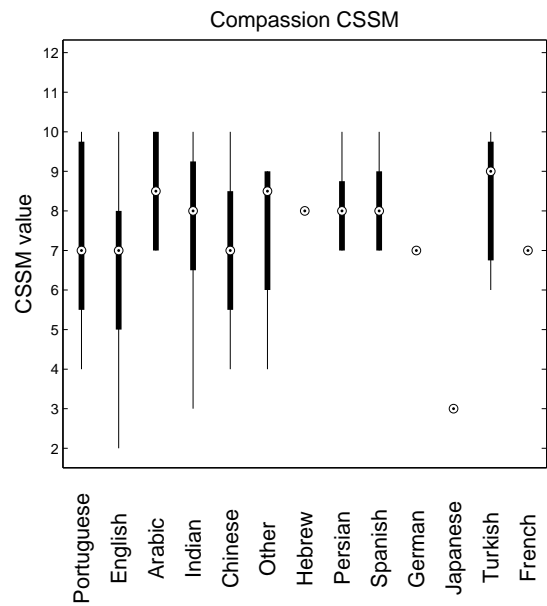
Figure B.17: Response to 'Compassion' CSSM where (a) CSSM(US, Compassion, Woman, User, User), (b) CSSM(US, Compassion, Vendor, User, User), (c) CSSM(US, Compassion, Vendor, Woman, User), (d) CSSM(US, Compassion, Woman, Vendor, User)

176
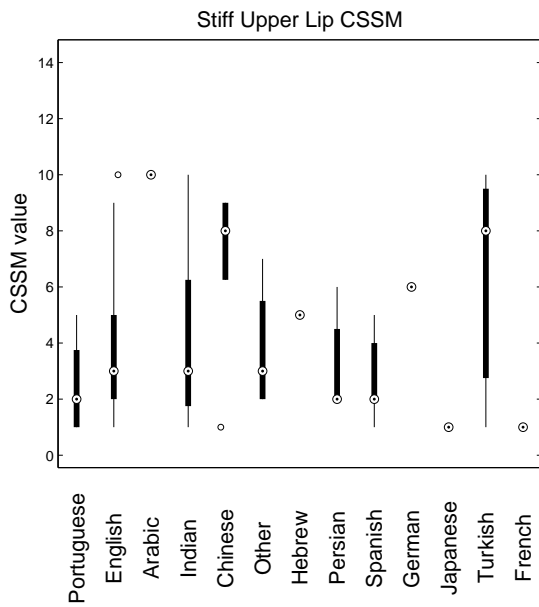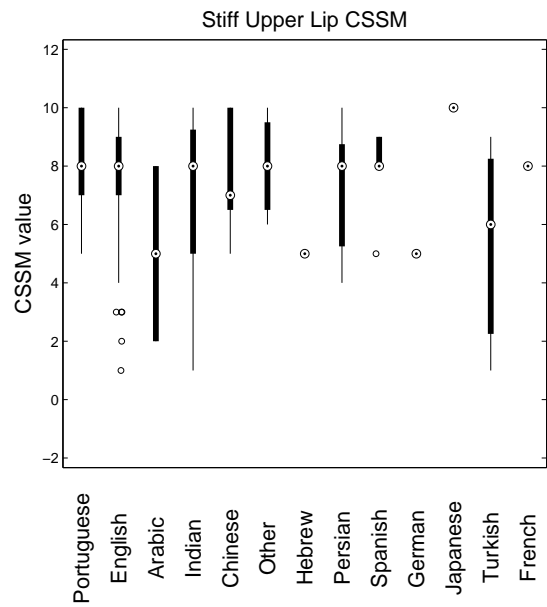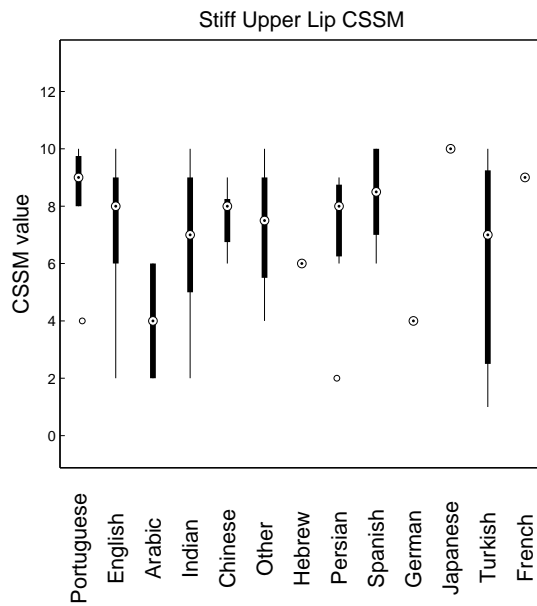
Figure B.18: Response to 'Stiff Upper Lip' CSSM where (a) CSSM(British, Stiff Upper Lip, Man, Woman, User), (b) CSSM(British, Stiff Upper Lip, Woman, Woman, User), (c) CSSM(British, Stiff Upper Lip, Woman, Man, User)

B.2   Excited Scenario

B.2.1   Overall response to CSSMs

(a)

(b)
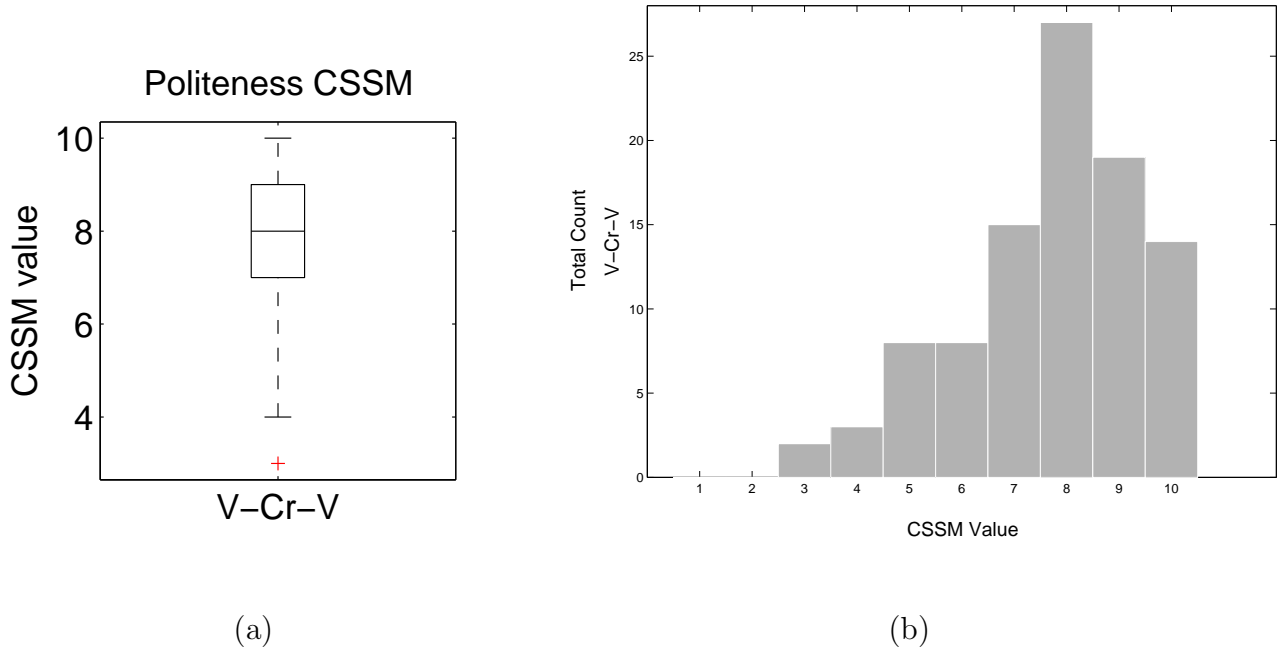
Figure B.19: Response to 'Politeness' CSSM where V-Cr-V = CSSM(Generic, Politeness, Vendor, Crowd, Vendor)
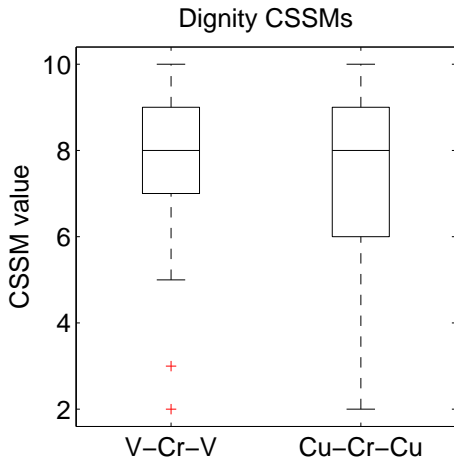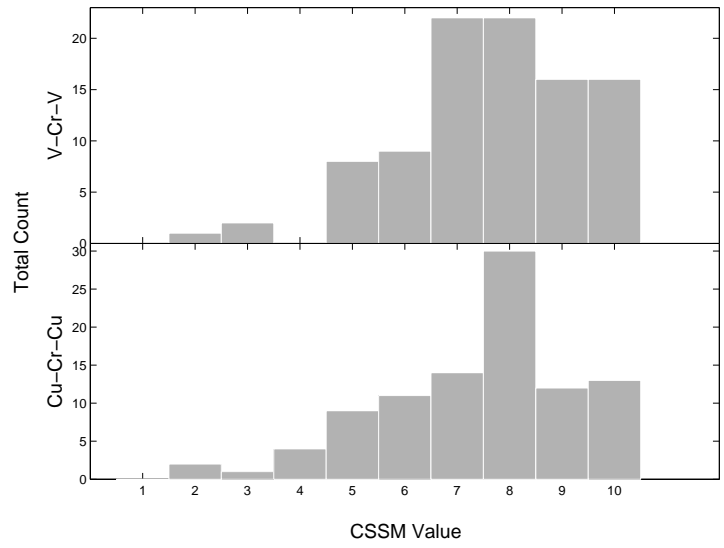
(a)   (b)

Figure B.20: Response to 'Dignity' CSSM where V-Cr-V = CSSM(Generic, Dignity, Vendor, Crowd, Vendor), Cu-Cr-Cu = CSSM(Generic, Dignity, Customer, Crowd, Customer)
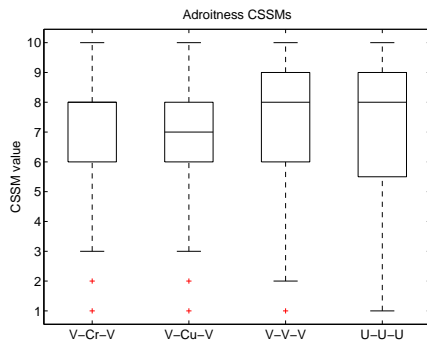
(a)
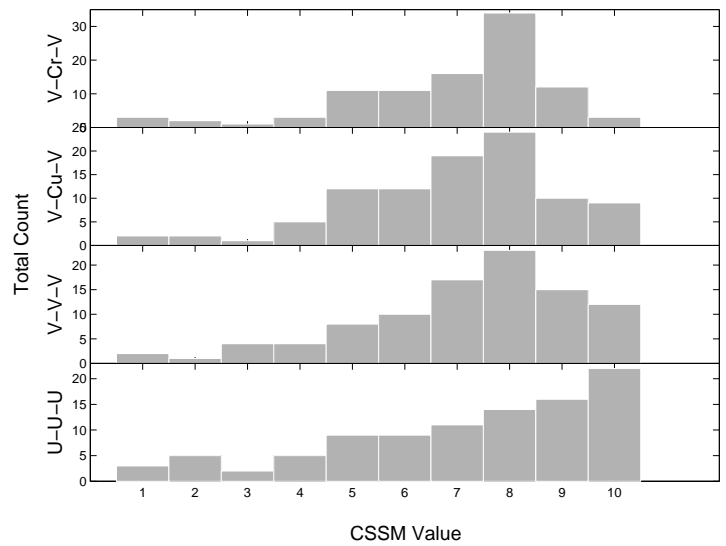
(b)

Figure B.21: Response to 'Adroitness' CSSM where V-Cr-V = CSSM(Generic, Adroitness, Vendor, Crowd, Vendor), V-Cu-V = CSSM(Generic, Adroitness, Vendor, Customer, Vendor), V-V-V = CSSM(Generic, Adroitness, Vendor, Vendor, Vendor), U-U-U = How well you understand the term "adroit"
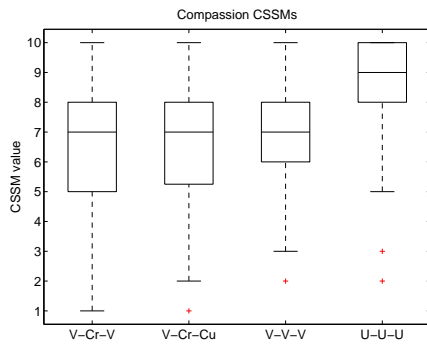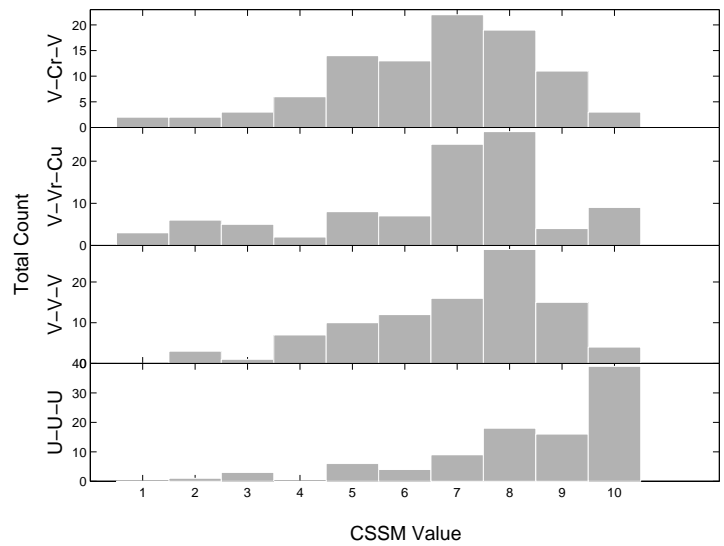
(a)   (b)

Figure B.22: Response to 'Compassion' CSSM where V-Cr-V = CSSM(Generic, Compassion, Vendor, Crowd, Vendor), V-Cr-Cu = CSSM(Generic, Compassion, Vendor, Crowd, Customer), V-V-V = CSSM(Generic, Compassion, Vendor, Vendor, Vendor), U-U-U = How well you understand the term "compassion"

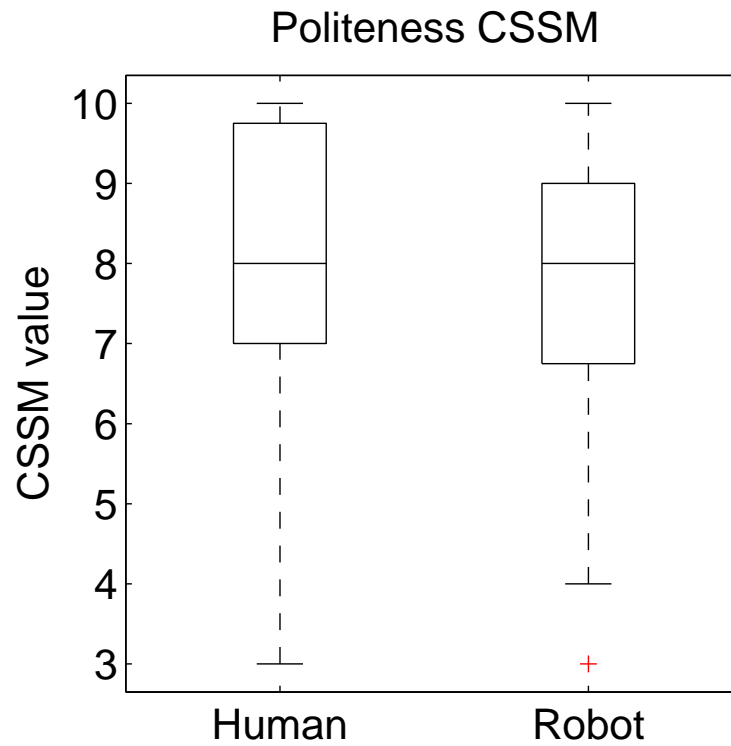## B.2.2 Comparison of Human to Robot scenario

Figure B.23: Response to 'Politeness' CSSM in Human vs Robot scenario, where CSSM(Generic, Politeness, Vendor, Crowd, Vendor)
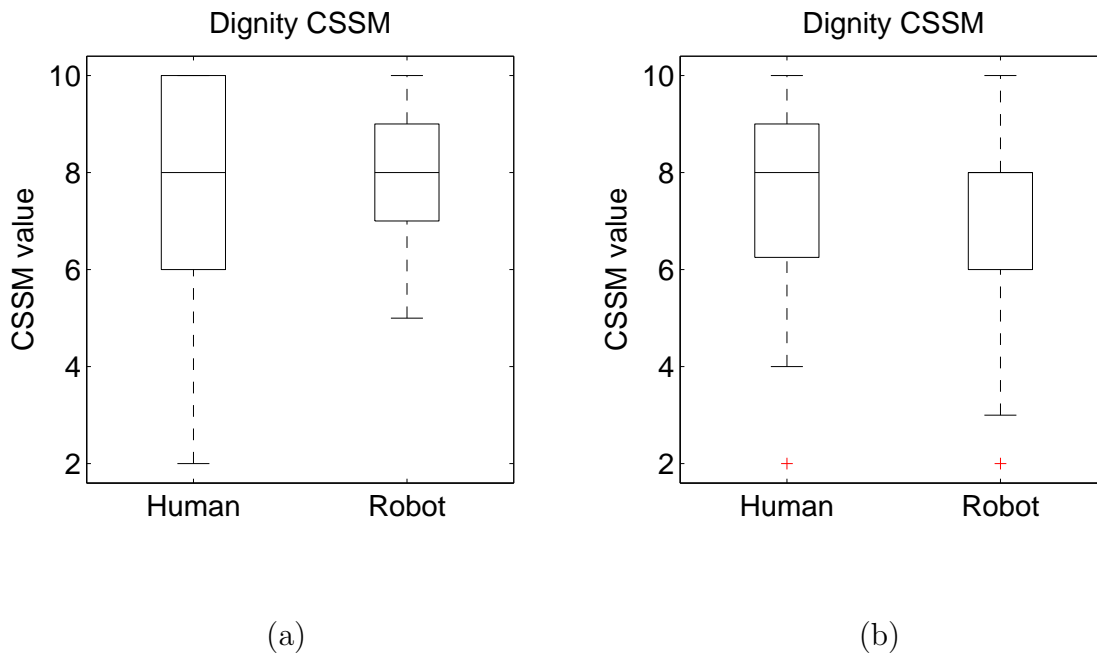
Figure B.24: Response to 'Dignity' CSSM in Human vs Robot scenario, where (a) CSSM(Generic, Dignity, Vendor, Crowd, Vendor), (b) CSSM(Generic, Dignity, Customer/Robot, Crowd, Customer/Robot)
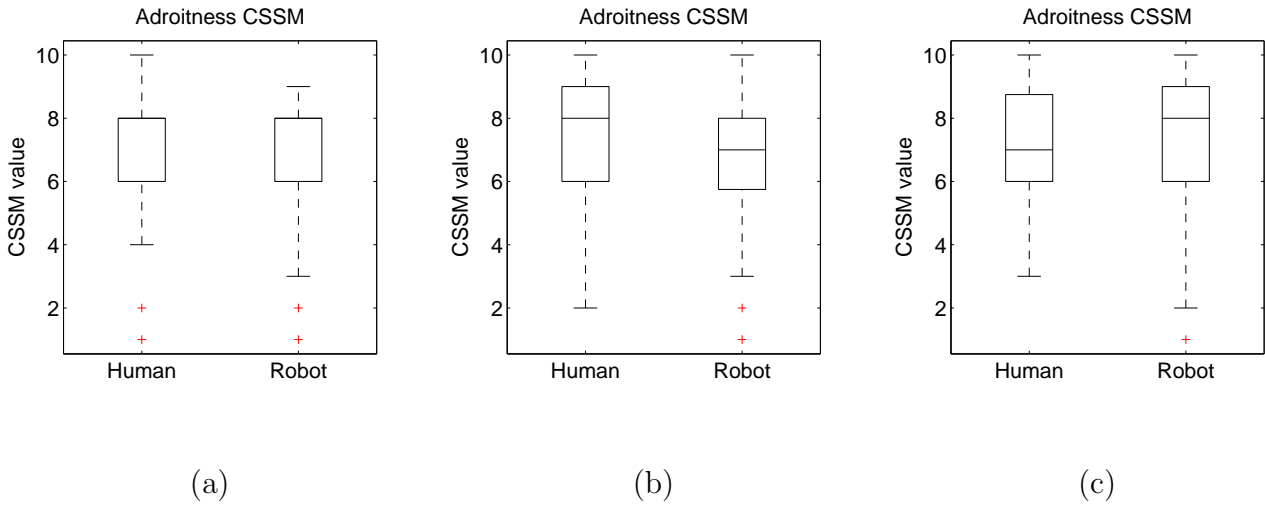
Figure B.25: Response to 'Adroitness' CSSM in Human vs Robot scenario, where (a) CSSM(Generic, Adroitness, Vendor, Crowd, Vendor), (b) CSSM(Generic, Adroitness, Vendor, Customer/Robot, Vendor), (c) CSSM(Generic, Adroitness, Vendor, Vendor, Vendor)
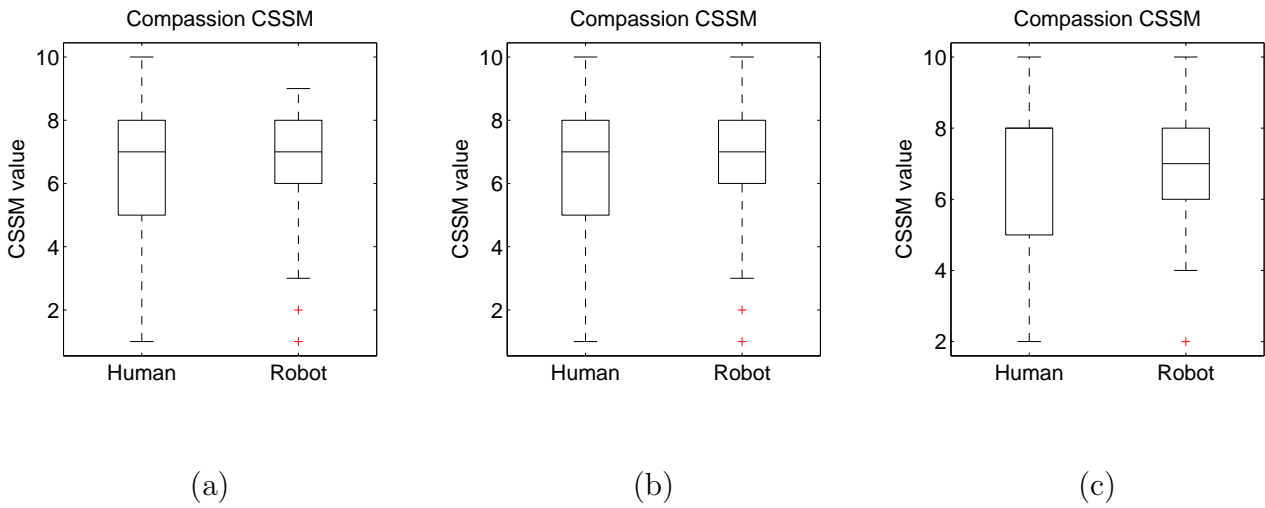
Figure B.26: Response to 'Compassion' CSSM in Human vs Robot scenario, where (a) CSSM(Generic, Compassion, Vendor, Crowd, Vendor), (b) CSSM(Generic, Compassion, Vendor, Crowd, Customer/Robot), (c) CSSM(Generic, Compassion, Vendor, Vendor, Vendor)

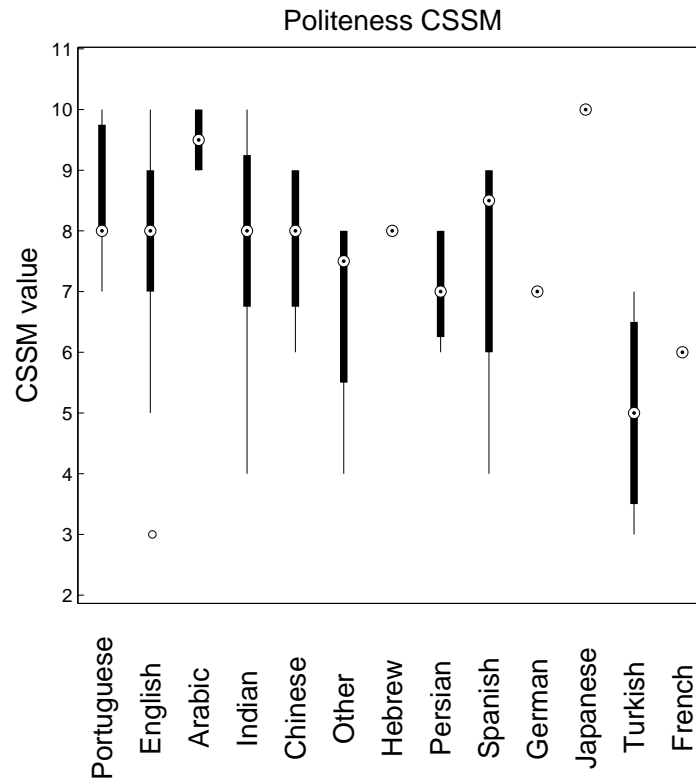## B.2.3 Comparison of CSSM evaluation categorized by participant language

Figure B.27: Response to CSSM(Generic, Politeness, Vendor, Crowd, Vendor)
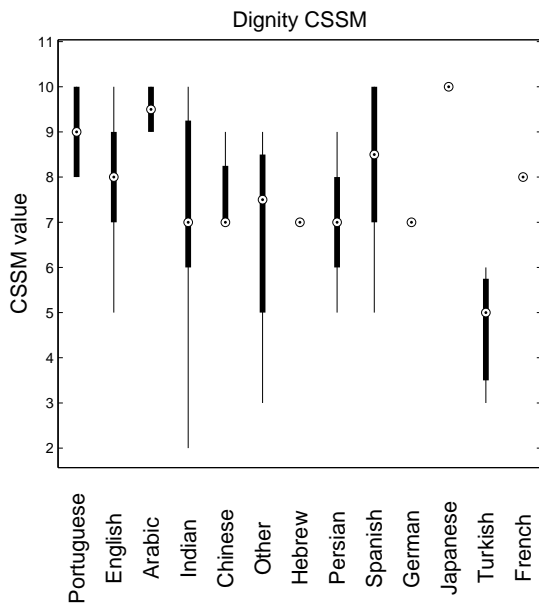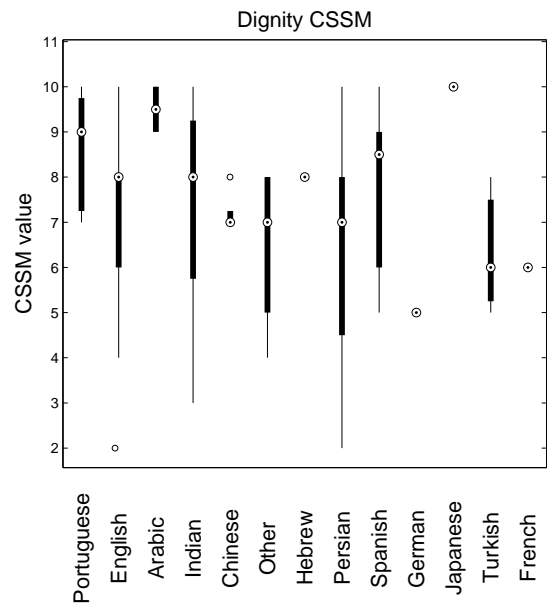
Figure B.28: Response to 'Dignity' CSSM where (a) CSSM(Generic, Dignity, Vendor, Crowd, Vendor), (b)CSSM(Generic, Dignity, Customer, Crowd, Customer)
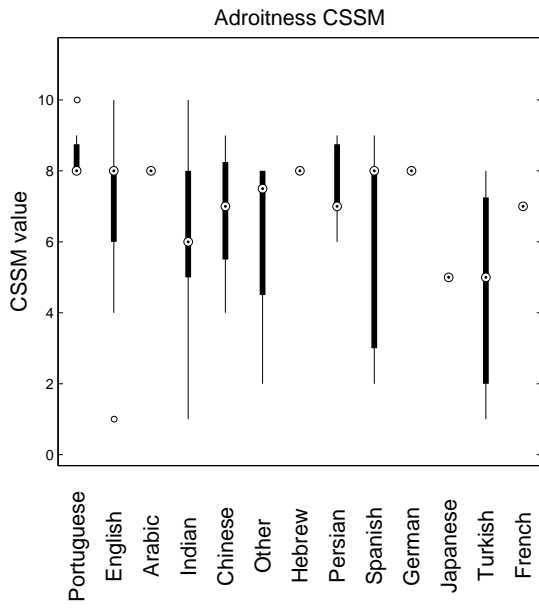
(a)
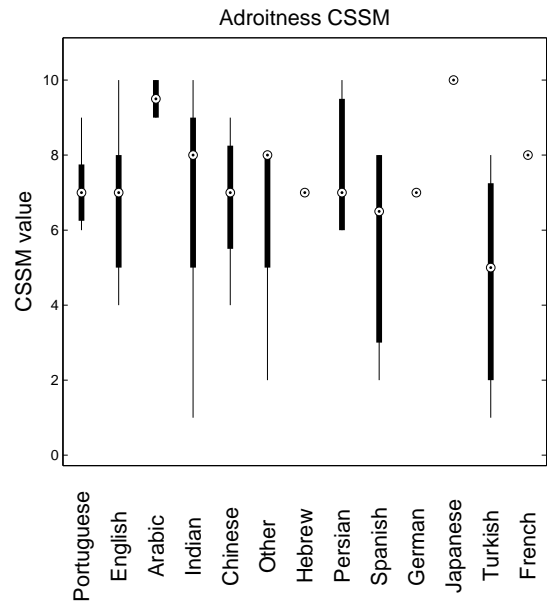
(b)
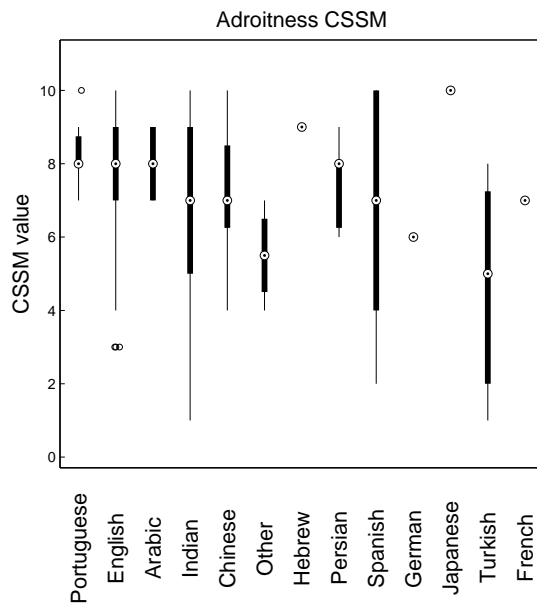
(c)

Figure B.29: Response to 'Adroitness' CSSM where (a) CSSM(Generic, Adroitness, Vendor, Crowd, Vendor), (b) CSSM(Generic, Adroitness, Vendor, Customer, Vendor), (c) CSSM(Generic, Adroitness, Vendor, Vendor, Vendor)
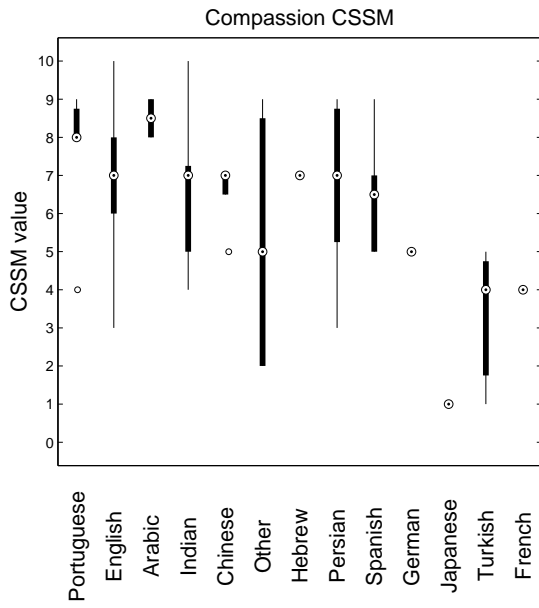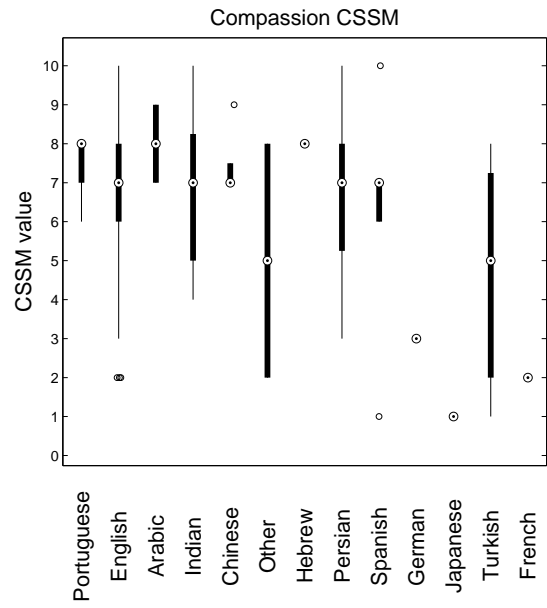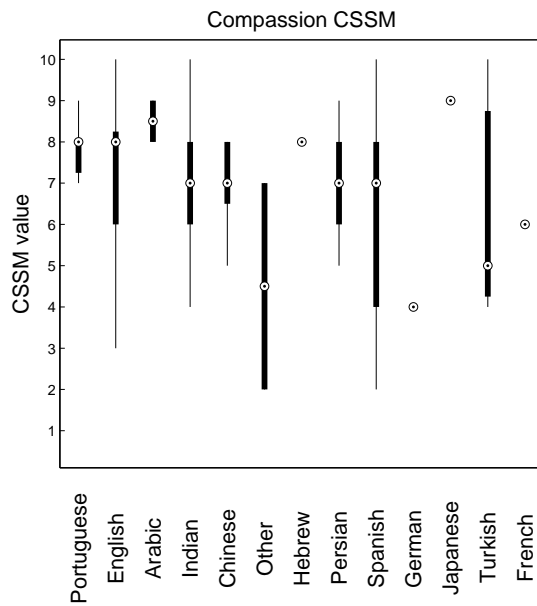
Figure B.30: Response to 'Compassion' CSSM where (a) CSSM(Generic, Compassion, Vendor, Crowd, Vendor), (b) CSSM(Generic, Compassion, Vendor, Crowd, Customer), (c) CSSM(Generic, Compassion, Vendor, Vendor, Vendor)

APPENDIX C
UCF IRB LETTER

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
             FWA00000351, IRB00001138

To: **Stephen M. Fiore** and Co-PIs: **Andrew P. Best, Katelynn A. Kapalo, Ladislau Boloni, Samantha Warta, Taranjeet Singh Bhatia**

Date: **May 17, 2016**

Dear Researcher:

On 05/17/2016, the IRB approved the following human participant research until 05/16/2017 inclusive:

Type of Review:	UCF Initial Review Submission Form
	Expedited Review Category #7
Project Title:	Investigating Social Cognitive Models Using Simulated Human Robot Interaction Scenarios
Investigator:	Stephen M Fiore
IRB Number:	SBE-16-12272
Funding Agency:	DOD/Army/ARL, General Dynamics
Grant Title:	RCTA-H9 Social Dynamics Theoretical Framework and Experimentation
Research ID:	1059059

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study.   All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 05/16/2017, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a signed and dated copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Page 1 of 2

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Stephen M. Fiore** and Co-PIs: **Andrew P. Best, Katelynn A. Kapalo, Ladislau Boloni, Samantha Warta, Taranjeet Singh Bhatia**

Date: **June 10, 2016**

Dear Researcher:

On 06/10/2016, the IRB approved the following minor modification to human participant research until 05/16/2017 inclusive:

Type of Review:	IRB Addendum and Modification Request Form
	Expedited Review Category #7
Modification Type:	Adding additional questionnaires to the study related to measuring social and cultural values. These questionnaires will be used in conjunction with the existing questionnaires to better understand some of our underlying research questions and they are also related to some of the overarching goals of the HRI Thrust within the RCTA project more broadly.
Project Title:	Investigating Social Cognitive Models Using Simulated Human Robot Interaction Scenarios
Investigator:	Stephen M Fiore
IRB Number:	SBE-16-12272
Funding Agency:	DOD/Army/ARL, General Dynamics
Grant Title:	
Research ID:	1059059

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 05/16/2017, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a signed and dated copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants

Page 1 of 2

192

LIST OF REFERENCES

[1] G. W. Liebniz, "The art of discovery," in *Liebniz: Selections* (P. P. Wiener, ed.), Charles Scribner's Sons, 1951.

[2] L. Bölöni, "The Spanish Steps flower scam - agent-based modeling of a complex social interaction," in *Proc. of 11th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, pp. 1345–1346, Jun 2012.

[3] T. Bhatia, S. Khan, and L. Bölöni, "The education of a crook: reinforcement learning in social-cultural settings," in *Proc. of Int'l Conf. on Autonomous Agents and Multi Agent Systems (AAMAS-2014)*, pp. 1397–1398, May 2014.

[4] T. Bhatia, S. Khan, and L. Bölöni, "A modeling framework for inter-cultural social interactions," in *Proc. of the Second Int. Workshop on Human-Agent Interaction Design and Models (HAIDM-13) at AAMAS-2013*, pp. 16–31, 2013.

[5] T. Bhatia, S. Khan, and L. Bölöni, "Towards an operational model for the propagation of public perception in multi-agent simulation," in *13th International Workshop on Multi-Agent Based Simulation (MABS-2012)*, pp. 1–12, June 2012.

[6] B. Edmonds and S. Moss, "From KISS to KIDS–an 'anti-simplistic' modelling approach," *Multi-Agent and Multi-Agent-Based Simulation*, pp. 130–144, 2005.

[7] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, no. 12, pp. 1743–1759, 2009.

[8] T. Parsons, *The social system*. Psychology Press, 1951.

[9] C. W. Mills, *The sociological imagination*. Oxford University Press, 1959.

[10] P. Brown and S. Levinson, *Politeness: Some universals in language usage.* Studies in Interactional Sociolinguistics vol. 4, Cambridge Univ Press, 1987.

[11] H. P. Grice, "Logic and conversation," in *Syntax and Semantics, Volume 3: Speech Acts*, pp. 41–58, Academic Press, 1975.

[12] Y. Gu, "Politeness phenomena in modern Chinese," *Journal of Pragmatics*, vol. 14, no. 2, pp. 237–257, 1990.

[13] Y. Matsumoto, "Reexamination of the universality of face: Politeness phenomena in Japanese," *Journal of Pragmatics*, vol. 12, no. 4, pp. 403–426, 1988.

[14] G. Hofstede, G. Hofstede, and M. Minkov, *Cultures and organizations: software for the mind.* McGraw-Hill Professional, 2010.

[15] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[16] R. Albert and A. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[17] J. Kottonau and C. Pahl-Wostl, "Simulating political attitudes and voting behavior," *Journal of Artificial Societies and Social Simulation*, vol. 7, no. 4, 2004.

[18] M. Motani, V. Srinivasan, and P. Nuggehalli, "PeopleNet: Engineering a wireless virtual social network," in *Proc. of Int'l. Conf. on Mobile Computing and Networking (MobiCom)*, 2005.

[19] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in *Proc. of the 13th Int'l. World-Wide Web Conference*, 2004.

[20] E. Adar and L. Adamic, "Tracking information epidemics in blogspace," in *Proc. of 2005 IEEE/WIC/ACM Int'l Conf. on Web Intelligence*, pp. 207–214, IEEE, 2005.

[21] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummad, "Measuring user influence on Twitter: The million follower fallacy," in *Proc. of the Fourth Int'l AAAI Conference on Weblogs and Social Media*, pp. 10–17, 2010.

[22] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.

[23] Z. Liu and B. Hu, "Epidemic spreading in community networks," *Europhysics Letters*, vol. 72, no. 2, p. 315, 2005.

[24] R. May and A. L. Lloyd, "Infection dynamics on scale-free networks," *Physical Review E*, vol. 64, no. 6, p. 66112, 2001.

[25] P. Bramsen, M. Escobar-Molano, A. Patel, and R. Alonso, "Extracting social power relationships from natural language.," in *Proc. of the Association for Computational Linguistics*, pp. 773–782, 2011.

[26] C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, and C. Potts, "A computational approach to politeness with application to social factors," in *Proc. of the Association for Computational Linguistics*, 2013.

[27] C. Miller, P. Wu, and H. Funk, "A computational approach to etiquette: operational-izing Brown and Levinson's politeness model," *IEEE Intelligent Systems*, pp. 28–35, 2008.

[28] C. Miller, P. Wu, V. Vakili, T. Ott, and K. Smith, "Culture, politeness and directive compliance," in *Proc. of the 5th Int'l. Conf. on Universal Access in Human-Computer Interaction*, pp. 568–577, July 2009.

[29] T. Bosse, C. Jonker, and J. Treur, "Formalisation of Damasio's theory of emotion, feeling and core consciousness," *Consciousness and Cognition*, vol. 17, no. 1, pp. 94–113, 2008.

[30] A. R. Damasio, *The feeling of what happens: Body and emotion in the making of consciousness.* Houghton Mifflin Harcourt, 1999.

[31] A. Paiva, J. Dias, D. Sobral, and R. Aylett, "Caring for agents and agents that care: Building empathic relations with synthetic agents," in *Proc. of Autonomous Agents and Multi Agent Systems*, pp. 194–201, 2004.

[32] L. Hall, A. Woods, K. Dautenhahn, D. Sobral, A. Paiva, D. Wolke, and L. Newall, "Designing empathic agents: Adults versus kids," in *Proc. of Intelligent Tutoring Systems*, pp. 175–196, 2007.

[33] A. Ortony, G. Clore, and A. Collins, *The Cognitive Structure of Emotions.* Cambridge University Press, 1998.

[34] S. Rodrigues, S. Mascarenhas, J. Dias, and A. Paiva, "I can feel it too!: Emergent empathic reactions between synthetic characters," in *Proc. of Int'l. Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2009.

[35] S. D. Preston and F. De Waal, "Empathy: Its ultimate and proximate bases," *Behavioral and Brain Sciences*, vol. 25, no. 01, pp. 1–20, 2002.

[36] F. De Vignemont and T. Singer, "The empathic brain: how, when and why?," *Trends in Cognitive Sciences*, vol. 10, no. 10, pp. 435–441, 2006.

[37] L. Hogg and N. R. Jennings, "Socially intelligent reasoning for autonomous agents," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 31, no. 5, pp. 381–393, 2001.

[38] J. C. Harsanyi, "Morality and the theory of rational behavior," *Social Research*, pp. 623–656, 1977.

[39] S. Kalenka and N. R. Jennings, "Socially responsible decision making by autonomous agents," in *Cognition, Agency and Rationality*, pp. 135–149, Springer, 1999.

[40] C. Castelfranchi, "Commitments: From individual intentions to groups and organizations," in *Proc. of the First Int'l Conf. on Multiagent Systems*, pp. 41–48, 1995.

[41] F. Grimaldo, M. Lozano, F. Barber, and G. Vigueras, "Simulating socially intelligent agents in semantic virtual environments," *Knowledge Engineering Review*, vol. 23, no. 4, p. 369, 2008.

[42] L. Bölöni, S. Khan, and S. Arif, "Robots in crowds - being useful while staying out of trouble," in *Proc. of Intelligent Robotic Systems Workshop (IRS-2013)*, pp. 2–7, 2013.

[43] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: a survey," *Foundations and Trends in Human-Computer Interaction*, vol. 1, no. 3, pp. 203–275, 2007.

[44] K. Dautenhahn, "Socially intelligent robots: dimensions of human–robot interaction," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, pp. 679–704, 2007.

[45] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson, and H. Yan, "Embodiment in conversational interfaces: Rea," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 520–527, ACM, 1999.

[46] C. Breazeal, "Emotion and sociable humanoid robots," *International Journal of Human-Computer Studies*, vol. 59, no. 1, pp. 119–155, 2003.

[47] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

[48] R. Nisbett and D. Cohen, *Culture of honor: The psychology of violence in the South.* Westview Press, 1996.

[49] D. J. Kiesler, "The 1982 interpersonal circle: A taxonomy for complementarity in human transactions," *Psychological Review*, vol. 90, no. 3, p. 185, 1983.

[50] L. E. Alden, J. S. Wiggins, and A. L. Pincus, "Construction of circumplex scales for the inventory of interpersonal problems," *Journal of Personality Assessment*, vol. 55, no. 3-4, pp. 521–536, 1990.

[51] D. Lerman, "Consumer politeness and complaining behavior," *Journal of Services Marketing*, vol. 20, no. 2, pp. 92–100, 2006.

[52] G. Shafer, *A mathematical theory of evidence.* Princeton University Press, 1976.

[53] R. Yager, "On the Dempster-Shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93–137, 1987.

[54] R. F. Baumeister, E. Bratslavsky, M. Muraven, and D. M. Tice, "Ego depletion: Is the active self a limited resource?," *Journal of Personality and Social Psychology*, vol. 74, no. 5, p. 1252, 1998.

[55] M. T. Gailliot and R. F. Baumeister, "The physiology of willpower: Linking blood glucose to self-control," *Personality and Social Psychology Review*, vol. 11, no. 4, pp. 303–327, 2007.

[56] F. Richards, "A flexible growth function for empirical use," *Journal of Experimental Botany*, vol. 10, no. 2, pp. 290–301, 1959.

[57] E. J. Lobato, S. F. Warta, T. J. Wiltshire, and S. M. Fiore, "Varying social cue constellations results in different attributed social signals in a simulated surveillance task," in *Proc. of 28th International FLAIRS Conference*, 2015.

[58] J. Henrich, S. Heine, and A. Norenzayan, "The weirdest people in the world," *Behavioral and Brain Sciences*, vol. 33, no. 2-3, pp. 61–83, 2010.

[59] T. Bhatia, G. Solmaz, D. Turgut, and L. Bölöni, "Controlling the movement of robotic bodyguards for maximal physical protection," in *Proc of the 29th International FLAIRS Conference*, pp. 380–385, May 2016.