UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Persuasive Social Agents using Social Power Dynamics

Mojgan Hashemian

Supervisor: Doctor Rui Filipe Fernandes Prada
Co-supervisor: Doctor Pedro Alexandre Simões dos Santos

Thesis approved in public session to obtain the PhD Degree in
Computer Science and Engineering

Jury final classification: Pass with Distinction

2021

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Persuasive Social Agents using Social Power Dynamics

Mojgan Hashemian

Supervisor: Doctor Rui Filipe Fernandes Prada
Co-supervisor: Doctor Pedro Alexandre Simões dos Santos

Thesis approved in public session to obtain the PhD Degree in
Computer Science and Engineering

Jury final classification: Pass with Distinction

Jury

Chairperson: Doctor Joaquim Armando Pires Jorge, Instituto Superior Técnico, Universidade de Lisboa

Members of the Committee:

Doctor James Everett Young, Faculty of Science, University of Manitoba, Canada

Doctor Luís Paulo Gonçalves dos Reis, Faculdade de Engenharia, Universidade do Porto

Doctor Rui Filipe Fernandes Prada, Instituto Superior Técnico, Universidade de Lisboa

Doctor Francisco António Chaves de Melo, Instituto Superior Técnico, Universidade de Lisboa

Doctor Rodrigo Martins de Matos Ventura, Instituto Superior Técnico, Universidade de Lisboa

Funding Institutions

European Union's Horizon 2020

Fundação para a Ciência e a Tecnologia

INESC-ID

2021

In memory of Zahra, Mohammad and all the victims of flight PS752...

Acknowledgments

This dissertation is dedicated to my family who have always been my biggest support and encouragement: To my parents, Mehrangiz and Hassan, without whose loving support this would not have happened. To my beloved sisters, for the unwavering support given at every step of the way, including my PhD and my whole life. And finally, a heartfelt thanks to the love of my life, Koorosh, for his unconditional patience and support all along.

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2019 and through the AMIGOS project (PTDC/EEISII/7174/2014), and the research project RAGE (Realising an Applied Gaming Eco-System) funded by the EU under the H2020-ICT-2014-1 program with grant agreement No 644187.

I am sincerely grateful to my supervisors, Rui and Pedro, for making this thesis possible with their input and patience. I extend my gratitude to Prof. Ana Paiva, director of the GAIPS lab, who always had a word of encouragement and motivation. And my special thanks to Samuel and Marta for their technical advice and aids. I am also grateful to Sandra who helped me dealing with the bureaucracies in Portugal.

I would like to express particular appreciation to my friends as well. To my dearest Zohreh who has always taught me about perseverance with her unexampled patience and admirable attitude. To Maryam, with whom I have shared moments of deep anxiety but also of big excitement. And to Sonia, for her endless support and kindness.

A very special word of thanks goes for my great friends Shruti, Marta, Raul, and Maria José, with whom I had the best tea breaks in my life. Their presence was essential in a process that is often felt as tremendously solitaire. A warm word to Elmira for all the moments we shared. I am also grateful to Filipa, Raquel and Carla, for valuable scientific discussions and useful suggestions. I would also like to express my deepest thanks to all my laboratory mates, for sharing many interesting discussions and work time fun. And many thanks to all volunteers who were willing to participate in the user-studies.

My acknowledgement would be incomplete without paying thanks to Dr Moradi, my Master thesis adviser, who introduced me to the interesting world of research. And I must add that the list is incomplete. I thank one and all for their blessings and support throughout my life.

Resumo

O poder social, o potencial de influência social, é um processo social generalizado nas interações humano-humano. Apesar de seu papel reconhecido na interação social, poucas atenções foram dadas a esse fenómeno na interação homem-agente. Um exemplo proeminente de agentes sociais que tem tido interesse nos dias de hoje é a evolução dos agentes robóticos. Além disso, em relação aos agentes robóticos, poucos estudos abordaram o poder social nas interações humano-robô.

Os recentes avanços na robótica social levantam a questão de saber se um robô social pode ser usado como um agente persuasivo. Até ao momento, foram realizadas várias tentativas usando diferentes abordagens para explorar esta questão de investigação. Em poucas palavras, o objetivo deste trabalho é dotar agentes inteligentes com dinâmica de poder social para desenvolver agentes mais racionais e, portanto, mais persuasivos. Por um lado, a adequada tomada de decisões do poder social aumenta a racionalidade dos agentes e, portanto, aumenta sua credibilidade. Por outro lado, possuir fontes de poder social aumenta a sua persuasão. Neste texto, reportamos os resultados de nossos avanços recentes para esse objetivo e levantamos questões sobre direções futuras.

Resumidamente, primeiro abordámos o problema conceptualizando o poder social inspirado numa teoria proposta por French e Raven. Em seguida, realizámos diferentes estudos com utilizadores para investigar de que forma diferentes bases de poder social contribuem para a persuasão dos robôs sociais. No Estudo 1, num cenário adversarial, programámos dois robôs com poder social pela perícia e pela recompensa. Como exemplo específico de recompensa, usámos recompensas sociais (contando uma piada pelo robô). No segundo estudo, usámos um único robô para convencer o utilizador através de duas estratégias diferentes (recompensa e coerção) em comparação com uma condição de controlo. Finalmente, no último estudo (Estudo 3), o robô utilizou uma estratégia de poder (recompensa), mas com níveis diferentes em condições diferentes, comparando-se a duas condições de controlo com / sem a presença de qualquer robô.

Em suma, os resultados de nossos estudos com utilizadores confirmam que o poder social (especificamente recompensa, coerção e perícia) confere persuasão aos robôs sociais. E diferentes estratégias persuasivas podem ser percebidas e preferidas de maneira diferente, considerando o perfil dos utilizadores ou as condições do estudo.

Palavras-chave: poder social, persuasão, robôs persuasivos, robôs sociais, agente socialmente inteligente

Abstract

Social Power, the potential for social influence, is a pervasive social process in human-human interactions. Despite its acknowledged role in social interaction, little attentions have been paid to this phenomenon in human-agent interaction. One prominent example of social agents that have been of interest these days is the evolution of robotic agents. Further, with respect to robotic agents, limited studies have addressed social power in human-robot interactions.

Recent advances on Social Robotics raise the question of whether a social robot can be used as a persuasive agent. To date, different attempts have been performed using several approaches to tackle this research question. In a nutshell, the objective of the present work is empowering intelligent agents with social power dynamics to develop more rational agents and hence more persuasive. On the one hand, proper decision making facing social power increases the rationality of agents and hence enhances their believability. On the other hand, possessing sources of social power boosts their persuasibility. In this text, we report the results of our recent advancements for this objective and draw suggestions for future directions.

In sum, we first approached the problem by conceptualizing social power inspired by a theory proposed by French and Raven. Then we performed different user-studies to investigate how different bases of social power contribute to persuasiveness of social robots. In Study 1, within an adversarial setting, we programmed two robots with expert and reward social power. In particular, we programmed one robot to express expertise in its behaviour by giving information to the users. Also, as a specific instance of reward, we used social rewards (telling a joke by the other robot). In the second study, a single robot was used to persuade the user using two different strategies (reward and coercion) comparing to a control condition. Finally, in the last study (Study 3), the robot used one power strategy (reward) but with different levels in different conditions, comparing to two control conditions with/without the presence of any robot. In this study, the persuasion attempt was repeated over a series of repeated interactions.

Overall, the results of our user studies endorse that social power (in particular reward, coercion, and expert bases) endows persuasiveness to social robots. And different persuasive strategies could be perceived and be preferred differently considering users' profiles or the study conditions.

Keywords: Social Power, Persuasion, Persuasive Robots, Social Robots, Socially Intelligent Agent

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> **The fundamental concept in social science is Power, in the same sense that Energy is the fundamental concept in physics.** *Bertrand Russell, 1938, p. 10*

1.1 Motivation

Social power is defined as the ability to influence other's attitudes, behavior and beliefs in an intended direction that may not happen in the absence of such ability [54]. This pervasive feature of our daily life has been proved to have a significant impact on social interactions. Previous research has established that social power is present in every relationship [51] and hence the rules of social dynamics can be stated in terms of power [150]. In other words, almost any relationship can be characterized in terms of the amount of control that an individual has over others' and vice versa [93]. In this sense, any relationship has a form of inequality due to power imbalances and hence causing social forces and dynamics.

Recent evidence suggests that social power (or for short power) is central to a multitude of social processes. Power acts as a *heuristic* solution to potential conflicts between group members and guides social perception and behavior [95]. Extensive research in the field of social psychology has shown that social power affects a wide variety of social and cognitive processes, such as stereotyping [48], moral judgement [47], as well as, nonverbal behavior, such as emotional display [32], and its inferences [65].

Altogether, the preceding paragraphs depict the importance of social power in social interactions. And to date, due to its importance, social power has been investigated in a vast variety of fields, such as politics, social leaders, religious leaders, organizations [158]. Here, we investigate social power from the perspective of computer science, specifically, Socially Intelligent Agents (SIA).

Additionally, recent evidence suggests that humans perceive computers as social agents and people respond socially to computer actors (Computers are Social Actors (CASA) paradigm). In other words, humans treat computers in a manner similar to how they treat other humans [121]. In this sense, people apply similar social rules to their relationship with computers [139]. The same might apply to social power theories with respect to computers. However, despite the important role of social power in interaction, to date, there are few studies that have investigated it in agent studies.

This research gap has motivated us to investigate social power in social agents. We approach this gap from two perspectives. Specifically, any power related relationship deals with two sides, an agent who exerts power (the actor) and the target. From the target side, we investigate decision making process within power related interactions. And from the actor side, we investigate his/her perception from the target side. Hence, to approach this research gap, initially we start by a conceptualization of power for social agent architecture. Based on this conceptualization we then propose different approaches on how to develop high-power agents and investigate their perceptions when using different power resources.

To be more specific, in this thesis, we aim to design agents capable of processing social power dynamics (a), as well as representing power in their behavior (b) and investigate how they are perceived having different power sources (c). To operationalize the expression of power sources, we propose to utilize persuasion, as an application of social power. Also, as a special type of social agents, we selected social robot due to their physicality that enhances the social presence (Section 2.4.1). In sum, in this dissertation, we aim to investigate the link between social power and persuasion in social robots. Specifically, we investigate how to design more persuasive robots using social power. And by operationalizing social power in context of persuasion, we develop different persuasion studies based on three different power bases. We investigated the effectiveness of these persuasion strategies by designing and implementing three different user studies.

1.2 Problem Statement and Research Goals

This study aims to contribute to the growing area of SIA research by exploring social power. The fundamental problem we address is how social power guides social interaction of intelligent agents. The answer to this research question is two-folded, depending on the perspective of each side of the social interaction. Hence, we divide this research questions into three research goals:

1. To conceptualize social power with less complexities (in comparison to a recent study) to have power-aware agents leading to social reasoning and decision making facing social power.

2. To investigate how social power makes the robots more persuasive.

3. To investigate how different sources of social power lead to different perception.

With this aim, we focus on theoretical issues for agent modeling in presence of social power. To do so, we have selected one of the most sophisticated models in this field and then we suggest some improvements (further details in Section 4.1). To be more specific, we extend a previously proposed model of social power (SAPIENT), which is based on the well-known theory of social power proposed by French and Raven [54]. By this extension, we decrease the complexity of SAPIENT and propose a model that is purely driven from the social power model proposed in [54].

Then to address the perspective of the target, we propose to use different power bases inspired by this model to equip the agents with the conceptualized power resources. As one key factor influencing

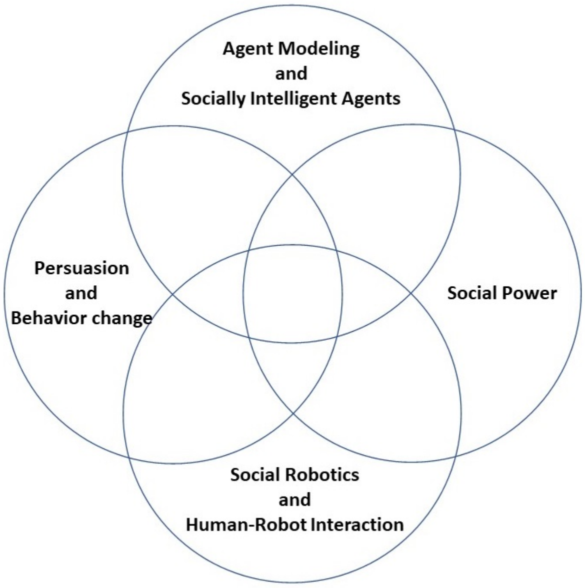


Figure 1.1: Research Context

the level of persuasion is the perception of the source power from the perspective of the target, by using different power strategies, we investigate how differently the persuader agents are perceived. To further measure the effectiveness of power resources, we use the concept of persuasion in different user-studies. Persuasion gives us the opportunity to operationalize social power within social interaction.

This thesis touches four different research domains as depicted in Figure 1.1. We do not consider these domains isolated from one another, rather to accomplish these research goals, we approach them as overlapping aspects.

1.3 Contributions

Our research led to contributions in the research area of socially intelligent agents. First, bearing in mind that social power plays a key role in the cognitive processes that mediate behavior, our conceptualization leads to agents that are more socially intelligent. Specifically, as social power is recognized as one of the factors which affects the decision making process, the model we propose leads the agents to make more rational decisions taking into account social power levels with less complexity compared to a previous study. Given the novelty of the approach, we published the following work in this direction:

- Hashemian, M., Prada, R., Santos, P. A., Mascarenhas, S. (2018, November). Enhancing social believability of virtual agents using social power dynamics. In Proceedings of the 18th International Conference on Intelligent Virtual Agents (pp. 147-152) [81].

Second, the presence of social robots in our daily lives creates new avenues for studying and developing persuasion strategies. In specific applications, robots are supposed to promote and/or encourage particular behaviors, or persuade a person to comply with a request or instruction to change (and/or) maintain a particular behavior. Hence, robots need to convey their persuasive strategies in a *socially acceptable manner* to gain higher behavior change. In this direction, we argue that using social power resources would enhance the persuasibility of these specific case of social agents. This part produced the following publications:

- Hashemian, M. (2019, May). Persuasive Social Robots using Social Power Dynamics. In 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS) (pp. 2408-2410) [70].

- M Hashemian, S Mascarenhas, M Couto, A Paiva, PA Santos, R Prada, The Application of Social Power in Persuasive Social Robots, Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot (HRI'2020) [75].

With this aim, we designed and implemented different user-studies to operationalize power resources in social robots. Our contribution in this direction is three-fold. First, we selected two of power bases (i.e., reward and expert) and programmed two robots in a competitive scenario to persuade the users to opt for a specific choice. Previous publication related to this aim includes:

- M Hashemian, A Paiva, S Mascarenhas, PA Santos, R Prada, Social power in human-robot interaction: Towards more persuasive robots, Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS) [78].

- M Hashemian, A Paiva, S Mascarenhas, PA Santos, R Prada, The Power to Persuade: A Study of Social Power in Human-Robot Interaction, The 28th IEEE International Conference on Robot & Human Interactive [77].

Second, we selected another base of power (i.e., coercion) and implemented another user study in a non-competitive setting. This study examined two bases of power, namely reward and coercion, in comparison to a control condition. The study and analysis of the findings resulted in the two following publications:

- M Hashemian, M Couto, S Mascarenhas, A Paiva, PA Santos, R Prada, Investigating Reward/Punishment Strategies in the Persuasiveness of Social Robots, Proceedings of the 29th IEEE International Conference on Robot and Human [71].

- M Hashemian, M Couto, S Mascarenhas, A Paiva, PA Santos, R Prada, Persuasive Social Robots using Reward/Coercion Strategies, Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot [72].

Third, within a game scenario, we investigated the influence of reward social power on persuasion. A key contribution of this part is the repeated persuasion attempts and the different levels of social power that the robot exerts over the participants. The result of this part is under publication (in Springer Persuasive Proceedings) and in review (IEEE Transactions on SMC).

- M Hashemian, M Couto, S Mascarenhas, A Paiva, PA Santos, R Prada, Persuasive Social Robot using Reward Power over Repeated Instances of Persuasion, Proceedings of the 16th International Conference on Persuasive Technology, 2021 [74].
- M Hashemian, M Couto, S Mascarenhas, A Paiva, PA Santos, R Prada, Building Persuasive Robots with Social Power Strategies, in IEEE Transactions on SMC: Systems (in review) [73].

The above mentioned contributions lead to different research impacts. We make the following contributions to the academic literature in SIA. First, one potential application of SIA is in the context of computer games, in which believable social interaction is a key requirement in order to provide a better gaming experience. It should be noted that, in modern computer games the need is not developing unbeatable games, but believable ones.

Second, rational decision making is a fundamental factor for the believability and effectiveness of interactive agent systems. We argue that representing the ability of reasoning and planning in the presence of social power enhances social believability of SIAs, leading to more sensible interactions. Particularly, believability of social interaction is not only about how the agents look “right” but also how they may do the right thing. Hence, to acquire more believable scenarios, nonverbal representation of power, in this case decision making, seems viable.

Third, designing powerful social robots contributes to having more persuasive robotics which is essential for a wide range of technologies, such as health care, energy saving, promoting physical activity, recruiting, etc. Also, recent studies indicate that persuasion might be used to encourage social interaction among older adults.

1.4 Other Contributions

Apart from the previously mentioned publications connected with this thesis, the list below represents other publications co-authored throughout the Ph.D. course.

- Do You Trust Me? Investigating the Formation of Trust in Social Robots, M Hashemian, R Paradeda, C Guerra, A Paiva, EPIA Conference on Artificial Intelligence, 357-369 [79].
- Inferring Emotions from Touching Patterns, M Hashemian, R Prada, PA Santos, J Dias, S Mascarenhas, 8th International Conference on Affective Computing & Intelligent Interaction (ACII 2019) [80].
- FIDES: How Emotions and Small Talks May Influence Trust in an Embodied vs. Non-embodied Robot, R Paradeda, M Hashemian, C Guerra, R Prada, J Dias, A Paiva, Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2017) [127].
- How facial expressions and small talk may influence trust in a robot, RB Paradeda, M Hashemian, RA Rodrigues, A Paiva, International Conference on Social Robotics (ICSR), 169-178 [129].
- “How is his/her mood”: A Question that a Companion Robot may be able to answer, M Hashemian, H Moradi, MS Mirian, International Conference on Social Robotics (ICSR), 274-284 [76].

- How facial expression may influence the trust in a robot?, RB Paradeda, M Hashemian, RA Rodrigues, A Paiva, 25TH IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) [128].

And below is the list of awards achieved during this Ph.D.:

- ACM Student Travel Grant for AAMAS 2019.
- AAAC Student Travel Grant for ACII 2019.
- HRI Pioneer: Selected to participate in the Human-Robot Interaction (HRI) Pioneers, a highly selective workshop to foster creativity, communication, and collaboration between young researchers, 2020.

1.5 Document Outline

This document has been organized in the following way. In Chapter 2, we begin by reviewing the background in the field of Social Psychology, i.e., current proposed models addressing social power. It will then go on with a brief review of the definition of persuasion in social psychology. And finally, we present an overview agent studies with respect to social power.

In the following chapter (Chapter 3) we review the state of the art in the field of Computer Science, i.e., recently developed models and architectures capable of representing or perceiving social power. Then we continue by discussing persuasive social robots.

The remaining part of the dissertation proceeds as follows: First, we give a brief overview of the proposed model in Chapter 4 and the upcoming chapters (Chapter 5-7) present the user-studies we performed. Then Chapter 8 begins by laying out the conclusion of the research, and continues by suggesting a set of potential steps for the future.

Chapter 2

Background

Research on social power has a long history and there has been a vast amount of literature addressing it mainly in the field of Social Psychology. To date, different theories exist in the literature regarding this concept. In this chapter, first, we review different theories of social power in the field of Social Psychology, and discuss how social power is related to persuasion, then we present an overview of the literature on persuasion. Finally, in Section 2.4, we briefly outline different types of social agents and the necessity of incorporating social power in their design. This chapter is concluded by a brief discussion in the last section.

2.1 Social Power

To date, different definitions of social power have been introduced in the field of Social Psychology. In this section, we review seminal works on social power in this field in chronological order. Based on the postulated theories in this field, social power (or for simplicity 'power') arises from static features of the individuals, such as their physical morphology, gender, ethnicity, etc. Apart from these static features, however, individuals can work to extend their level of power through other processes independent from their static features. In other words, people can promote their formal power, such as their status, social class, etc. by different social interactions with other members of society [95].

One seminal work in the field of Social Psychology addressing understanding of social power between humans is introduced by Dahl in reference [35]. In this work, social power was defined in terms of a relationship between people, as one's ability to influence the other one to do something which s/he would not do without presence of such power. From this view, people can be ranked based on their degree over a specific phenomenon of power that they possess. Based on this theory, social power has been conceptualized using three main factors: scope of power (factual and thematic range), number of comparable respondents (the social dimension abstracted into a mere number of subordinate), and changes in probabilities (the temporal dimension of the readiness to accept, convinced however not in terms of permanence but as change).

Another model of social power which is one of the most famous and highly refereed works on social power is performed by French and Raven [54]. In this work, the authors have identified different bases of power, including reward, coercion, legitimate, expert and referent (further discussed in Section 2.1.1). While there have been other identified bases, the authors argue that these five bases are the most common ones among the others [54]. In other words, most of other frameworks can be mapped onto these 5 bases. This model has been verified 30 years later [136].

Another well-known theory of social power was introduced by Emerson in [46]. Similarly, in this view power must exist in a dyadic relationship, i.e. a mutual dependence. However, in this work the focus of the model is on the type of relationships. The strength of this theory is its generality in covering a wide range of different social movements, such as internalization of social norms. In this theory, personal characteristics have not been considered as a factor in social power, since they are supposed to be highly variable across different relationships. On the other hand, Power has been defined in terms of dependence. The definition of dependence between two actors is based on two characteristics: The dependence of actor A upon actor B is (1) directly proportional to A's motivational investment in goals mediated by B, and (2) inversely proportional to the availability of those goals to A outside of the A-B relation."In this theory, when the two sides of a relationship need each other equally, their dependence is mutual. However, when there is no balance in the need, the side of the relationship who is more independent has more power. In other words, when A wants/needs something that B owns, this dependence relationship provides more power to B over A.

Another well-known theory of social power, i.e., the Approach/Inhibition Theory of Power is proposed in [94]. This model addresses the influence of power on an individual's behavior. This theory is based on the behavioral approach system (BAS) and behavioral inhibition system (BIS), that influence the sensitivity to reinforcement and punishment. These two brain-base systems control one's interaction with the environment. BAS regulates desirable motivations, and in this system the goal is to move toward something attractive. On the contrary, BIS regulates “aversive motives”, in other words in this system the goal is to move away from something unpleasant [63]. Most organisms have been shown to display one of these two types of reactions within the environment. In doing so, this theory of social power states that power has the ability to transform individuals' psychological states. In other words, power triggers activation of these two systems. Specifically, based on this theory, powerful individuals develop a cognitive bias that is more approach-related. Also, high power individuals are supposed to experience higher positive affects (such as pride) and process their life events in a more simplistic way. On the contrary, low power individuals develop inhibition-related tendencies over time. They experience more negative affects (such as shame), also they tend to behave more similarly to the others.

Another theory of social power was introduced in [114] which is mainly intended for organizations. This theory identifies four primary sources of power: ideological, economic, military, and political (IEMP). Each of these sources offers alternative organizational means of social control. Based on this theory, each of these sources creates independence while complementary sources of power for the individuals within each category.

Based on another definition, provided in [44], power involves getting individuals to comply with requests, even if they are reluctant to do so. It should be mentioned that, to date, several types of power have been recognized in the field of Social-Psychology, including organizational power, political power, etc. However, in this proposal, what is meant by power is the concept of social power, especially based on Raven and French definition in [54]. In this definition, power is defined in terms of influence, and influence is defined in terms of psychological change.

In sum, there has been a broad range of theories and definitions of social power based on the objective of the studies. The proposed models can be categorized based on different criteria. For instance, the recent models can be categorized in three different groups based on the way they define power [49]:

1. Power as influence: This category refers to the models which define power by its effect. Generally speaking, in these models when someone causes another to behave in a certain way, the former has power over the latter. (e.g. Emerson [46]).

2. Power as potential influence: These models emphasize the possibility to influence. In other words, the powerful person has the control over valued resources, and hence s/he has the potential to influence (e.g. French et. al. [54] and our perspective).

3. Power as outcome control: In this view, power is considered as a resource or outcome control. In other words, these models identify power by its consequences and hence defines power in terms of what it does, not what it is. So, the powerful agent has the control over others' valued outcome constituted power, even if the target resists (e.g. Dahl [35]).

From another perspective, models of social power can be categorized into four groups considering the scope of relationship and levels of analysis [40]: inter-personal, intra-personal, inter-group and ideological process. The first two types deal with individuals, inter-group deals with individuals who belong to the same group based on specific social identification, status, and different power positions. The last category focuses on how individuals construct social representation to legitimacy or how they consider their position in a social structure.

2.1.1 Discussion

The comprehensive concept of social power might lead to other categorizations. Among all other models and theories of social power, we have selected the model introduced by French and Raven [54] mainly due to the specific application of potential influence within interpersonal relationships. Another motivation for selecting this model is its generality in defining power sources. As mentioned earlier, the authors argue that these five bases are the most common ones among the others [54]. Also, the model has been verified after thirty years which establishes its reliability. The model is one of the most popular and its widely accepted conceptualization is examined in terms of applicability to various settings [44]. Also, the model is based on a typological analysis of the bases of power in interpersonal influence, which makes it very interesting for our second goal of designing persuasive agents. In this model, the authors have identified different bases of power, i.e., resources that can make changes in another person's belief, behavior, attitude (reward, coercion, legitimate, expert and referent). The definition of each five bases are as follows:

- **Reward** social power exists when the target is willing to do the requested action by the actor in response to another action which brings values to the target in his/her perspective. As an example, a factory manager [the actor] promises an employee [the target] to double his/her salary, if he/she increase production.

- **Coercive** power stems from the ability of one individual to mediate punishments for another. An example of this base of social power is as follows: a factory manager [the actor] tells a worker [the target] that if he does not increase production, then he will be fired.

- **Legitimate** power stems from internalized values that give one individual the authority to influence another. An example of this base of social power is as follows: in a family, a parent [the actor] instructs a teenage child [the target] to be home before midnight. In case of a successful interaction, the teenager gets home earlier because of the father's authority.

- **Expert** power stems from one individual's perception of others' higher knowledge. An example of this base of social power is as follows: a physician [the actor] instructs a patient [the target] to follow a given medical prescription.

- **Referent** power stems from the *identification* of one individual with another. An example of this base of social power is as follows: a person [the actor] asks a *friend* [the target] for help in studying for an upcoming exam.

Apart from the research performed in the field of Social Psychology, researchers in other fields such as Computer Science, Marketing, etc. have also investigated social power with different motivations. In Section 3.1, we briefly overview a number of recent literature performed in the field of Computer Science focusing on Modeling social power.

2.2 Persuasion and Power Relationship

Based on the definition of power, i.e. the ability to influence, the relationship between social power and influence is already established. On the other hand, persuasion is supposed as "an important medium of social power" [41]. This motivated us to investigate the effect of social power dynamics on persuasiveness of social agents and its potential effect on influencing the others. In this view, persuasion is defined as an attempt to change/shape a target's belief or behavior about a subject, an issue or an object [50] (further discussed in section 4.2).

The link between power and persuasion has been investigated in the field of social psychology for a long time [34] (for a recent review look at [18]). Early results show that a powerful individual is more influential in persuading others [86]. However, it should be noted that the extent to which that the power is effective is dependent on the circumstances by which it can cause short/long-term influence, as well as, to increase or decrease persuasion [18]. Specifically, some theories indicate a linear correlation between power and persuasion. In other words, in the higher power, the higher persuasion is achieved. However, recent evidence argues that this is generally not true. Under specific circumstances, when higher power is exerted, reactance happens and decreases the chance of persuasion. This happens due to the fact that persuasiveness of messages is dependent on the psychological sense of power. Hence, high power communicator may lead to high or low persuasion depending on the power state of the audience (the persuasion target).

For instance, evidence suggests that during mock interviews, when there exist a match between the power level of the interviewer and interviewees, higher persuasion is achieved [42]. In other words, high power communicators are more effective in persuading high power audiences, and similarly, low power communicators are more effective in persuading low power audiences. For instance, when both interviewer and interviewee are in low power state, the interviewer finds the target more persuasive. This is contradicting with earlier studies that stated interviewees with high-power are more persuasive [106]. Recent findings suggest that this inconsistency in the results is due to the mismatch between the powers, i.e. low-power interviewees were communicated by high-power interviewer. Hence, persuasiveness of messages is dependent on the psychological sense of power of the two sides.

Additionally, people with high power state, generate and pay greater emphasis on information that convey competence (e.g., stressing skillfulness and intelligence). On the contrary, low power state leads to more warmth, i.e., low-power communicators generate messages with more warmth information, for instance stressing friendliness and trustworthiness [42].

2.3 Persuasion

Having discussed the link between power and persuasion, it is necessary to define persuasion in further detail. In this section, we briefly review the proposed theories of persuasion in the field of social psychology. In general, persuasion is defined as an attempt to change/shape a target's belief or behavior about a subject, an issue or an object [50, 156]. Hence, persuasion involves the study of attitudes and how to change them [133]. In other words, persuasion may be defined as the formation of attitude change through information processing in response of a message about an intended object [13].

Within this process, by transmitting a message, the communicator tries to convince other people to change their own attitudes or behavior regarding an issue, in an atmosphere of *free choice*. It should be noted that the communicator does not change people's minds, but people decide to alter their own attitudes (in case of a successful persuasion) or to resist in front of the persuasion attempt (reactance) [133]. In this sense, *persuadability* is not supposed as an individual characteristic, but is a “complex communication phenomenon”.

Persuasion is a key process in shaping and maintaining cooperation, social influence and behavior change [138]. It plays a critical role in human interaction and exchanges [126] and a number of factors contribute to its effectiveness, such as the personality of the actor (the source or the one who is performing the influence) and the target (the one who is affected) [5, 126]. It should be noted that there is contradicting evidence regarding personality, for instance, some psychologists believe that due to the complexity of human behavior no personality trait is associated with persuasion [133].

To understand the process of *being persuaded*, the target's perception of the persuader's characteristics becomes important (for example, the internal cognitive process of the target). On the contrary, to understand the process of *persuading*, the characteristics of the actor play a vital role (e.g., actions of the actor).

The strength of influence is a function of the mode or the manner in which it is exerted [12], either loud, or forceful, threatening, sarcastic, softer, friendlier, light-humored mode, etc. Also, the persuasion message might vary in the degree to which it expresses such factors as hesitation, hedges, tag questions, and disclaimers. These differences in the message, might lead to powerless versus powerful speech [133].

Generally speaking, persuasion might happen through interpersonal interaction (interactive context), intra-personal communication and mass encounters (such as TV advertisements). Some scholars argue that persuasion involves awareness of the actor (persuader), in other words s/he is aware of her/his attempt to influence the target (persuadee). While others disagree with the awareness of the persuader (for instance [56]). Also, the target consciously/unconsciously decides to change his/her mind in that direction [133].

In sum, a simplified process of persuasion leads to a linear approach to the persuasion process [116], including the source of the persuasive message, the message, and the personality of the persuasive message receiver. However, since scholars have different perspectives on persuasion, hence different definitions and models of persuasion have been proposed to date [56]. In general, these theories can be categorized in the following groups [116]:

1. message effect theories (*)
2. cognitive processing theories (*)
3. functional approaches
4. attitude-behavior approach
5. consistency theories
6. Inoculation theory

Among the preceding categories, the first two (marked with *) lie in our interest and the scope of this thesis. In the reminder of this section, we present some of the most well-known theories of persuasion that lie within these two groups.

2.3.1 Message effect theories

This category of models focuses on the design of persuasive messages and provides concrete strategies to yield higher persuasion. From this perspective, a number of factors are recognized involved in persuasion process: 1. message structure (e.g., one-sided vs. two-sided messages) 2. types of evidence (e.g., vivid case history narratives), 3. fear appeals, 4. extended parallel process model [EPPM], 5. language use (e.g., speed, intensity).

One of the most renowned examples of this category is the model of persuasion that has been proposed in [56]. This model is based by limiting criteria for defining persuasion. Specifically, the authors have proposed five limiting criteria as follows: 1. intent criterion, 2. effect (has persuasion taken effect if no one is actually persuaded?), 3. free will and consciousness awareness (people are not always aware that persuasion is occurring), 4. symbolic action (often includes at least some coercion features), 5. inter-personal vs intra-personal (can persuasion happen between only two people or more than 2 distinct persons, as well as in self-persuasion.)

An interesting feature of this model is the distinction between intentional and unintentional persuasion. Particularly, based on this model, pure persuasion requires intention, and on the contrary, borderline persuasion does not. In other words, unintentional persuasion happens without any conscious awareness. For instance, consider a scenario in which two people are discussing betting on a specific horse, while a third person is listening to their conversation. Although the first person is trying to influence the second person, the third person might also be influenced; while s/he is not the subject of the persuasion attempt.

Some models consider intention directly in the definition of persuasion (for instance [133]). In their view, a persuader must intend to change another and deliberately attempt to influence. Hence, the persuaders must be aware that they are trying to achieve this goal. However, this model considers both pure (the core of persuasion) and border-line persuasion, with a fuzzy dividing border.

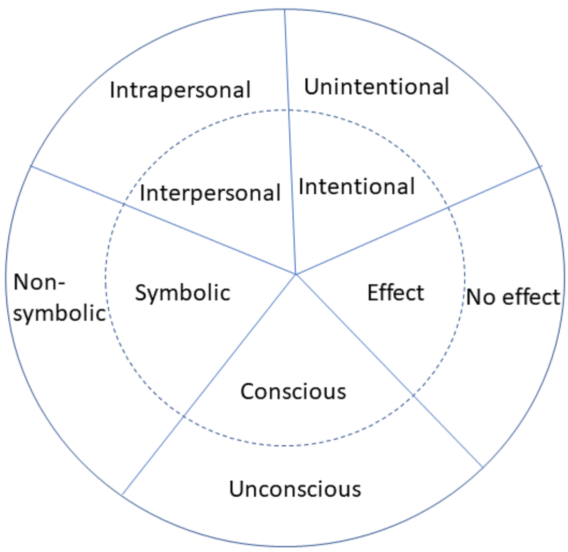


Figure 2.1: The model of the scope of persuasion

As depicted in Figure 2.1, the internal circle represents the *pure persuasion* and the outer circle represents the *border-line*. The dashed border represents the fuzzy distinction between the two cases. And the five wedges represent the five limiting criteria. It should be noted that, the five wedges can happen either in the intentional and unintentional persuasion.

Some theories of persuasion state that persuasion and coercion coexist. However, from the perspective of this model, persuasion requires an atmosphere of free choice. Specifically, coercion occurs when the influence agent delivers a believable threat of some consequence, that deprives the individual of some measure of freedom or autonomy, and attempts to induce the individual to act contrary to his or her preferences. On the other hand, persuasion occurs in an atmosphere of free choice, where the individual is autonomous, capable of saying no, and able to change his or her mind about the issue. In such situations, individuals are responsible for their choices and accountable for their decisions. [133]. Hence, in the view of this model, coercion and persuasion must not necessarily coexist.

Additionally, to reflect the nature of persuasion as a process, we need to consider the context in which the persuasion has happened. A number of factors specify the context of persuasion:

- The scope of the communication: either the persuasion happens in a 2-way interaction, or in a small group, or via mass media, or inside organization, and so forth.
- How synchronous or asynchronous is the communication, for instance, either it happens in a face to face communication, or there is a delay in between, such as sending emails, text.
- The ratio of verbal and non-verbal cues: type of media
- The goals of persuasion: a. self-representational, b. relational goals, c. instrumental goals
- socio-cultural factors: some cultures prefer more indirect approaches

Based on this theory, persuasion involves one or more persons who are engaged in the activity of creating, reinforcing, modifying, or extinguishing beliefs, attitudes, intentions, motivations, and/or behaviors within the constraints of a given communication context. In this definition, persuasion is not just changing their mind, it also can involve creating new beliefs or attitudes, where none existed before or strengthening or solidifying attitudes already held.

Another interesting theory of persuasion dealing the message effect is proposed by Cialdini in [31]. This model identifies six principles of persuasion as follows:

1. Principle of reciprocation: People feel obligated to return a favor.
2. Principle of scarcity: When something is scarce, people will value it more.
3. Principle of authority: When a request is made by a legitimate authority, people are inclined to follow/believe the request.
4. Principle of commitment and consistency: People do as they told they would.
5. Principle of consensus: People do as other people do.
6. Principle of liking: We say 'yes' to people we like.

These six principles, directly and indirectly correspond to the five bases of social power introduced by French and Raven in [54]. Specifically, reciprocation, authority, and commitment are dealt with legitimate power. Consensus principle, or conformity with majority opinion is sometimes based on a respect for the collective wisdom of the group, in which case it is expert power. And the rest, scarcity, and liking correspond to referent power.

2.3.2 Cognitive processing theories

In this section, we briefly review some of the most renowned models of persuasion dealing with cognition. We start by the ELM or Elaboration Likelihood Model [134]. This model proposes a framework for understanding the formation of attitude and behavior change in response to a persuasion attempt. Based on this model, persuasive messages are processed either centrally or peripherally. In other words, the persuader's characteristics can affect persuasion through different processes. Hence, there are two major routes to persuasion: central route in which the persuasive message is relevant to the persuadee and in this case the quality of the arguments have an influence on attitudes; and peripheral route in which the persuasive messages are less relevant to the persuadee and in this case the expertise of the source will be influential to the change in the attitudes.

In another model presented in [27], the authors propose that persuasive messages can be processed in two ways: heuristic and systematic information processing. Inspired by these two ways of information processing, which leads to different routes to persuasion, they have proposed a model of persuasion which is similar to the previous one that is composed of a dual-process system. Similarly, this model also focuses on how much likely are individuals to do thinking before making a decision. Specifically, in systematic route, information is processed in depth. In other words, individuals use the information to reach decision. On the contrary, in the heuristic approach, processing is done superficially. So, individuals make decisions faster using a rule of thumb and spend less cognitive effort. Hence, when it comes to making decision in response to a persuasive attempt, selecting each strategy depends on their motivation and their ability to process information at that current situation. For instance, in a case of lacking motivation or time for processing the information, heuristic information processing route is selected.

The last previous models focus on dual-process system for information processing, in other words how information processing behavior can lead to decision outcomes. The former focuses on the type of argumentation, while the latter deals with the depth of processing the information. There have been other models of persuasion which account for single-process or a "unimodel". For instance, in [103] the authors have approached persuasion from a single-route that treats the two former route equivalently. In this view, the two routes can occur simultaneously. In other words, when it comes to decision making in response to a persuasive attempt, both motivation and cognitive abilities are activated. In other words, there is "functional equivalence between cues/heuristics and messages/arguments in persuasion". Comparing to the previous two models, the unimodel has received greater recognition and acceptance in the literature in recent years [28].

2.3.3 Discussion

As discussed earlier, there have been different models of persuasion, all arguing that by a successful persuasion attempt, attitude change is formed [12]. These model either focus on the design of a persuasive message or the cognitive process of persuasion.

In general, we can categorize the models from different views: unimodel vs. dual-process model [12]. Earlier research on persuasion (in the 1980s and 1990s) was guided mainly by two dual-process models, the ELM [134] and the heuristic systematic model [27]. These models mainly focused on dual-process modes of information processing. Later, these models were challenged by the unimodels such as [12].

2.4 Social Power for Social Agents

Social power is one important attribute of the influencing agent in an interpersonal influence situation [88]. It generates psychological states which influence how we feel, think and act [6, 120]. Recent studies reveal that in the higher feeling of power, people's action orientation is greater, also power boosts the level of abstract thinking and leads to higher optimism in perceiving risk [120]. Moreover, feelings of power affect a wide range of interpersonal behaviors, such as verbal communication [154].

Also, the fact that people apply social rules to computers [122] highlights the need for capability of reasoning and processing, as well as, representing social phenomena for social agents. Specifically, when an agent appears sufficiently social, people communicate in a manner consistent with how they typically communicate in human–human interactions [102]. As social concepts contribute to social intelligence, social power enhances intelligence capability and believability [131].

On the other hand, in closer relationships the role of social power becomes a key factor. Specifically, in close relationships people depend upon each other to achieve their goals [43]. Social interaction is a core element of human experience [155]. In particular, social phenomena are inherently complex and require sophisticated mechanisms to express in automated form. While the social sciences provide inspiration in the form of explanatory analyses of social behavior, there is a need to transform those models into generative forms that can drive synthetic agents. With this aim, several research efforts have used sophisticated social models to support dynamic and believable interactions to develop different forms of social agent models.

Regardless of significant advances in intelligent agents' social capabilities over the last decade, there is still a significant gap in their social intelligence that seriously limits the range of social phenomena that can be simulated [131]. Although extensive research has been carried out with this aim, limited studies exists which investigate the role of social power in Socially Intelligent Agents.

2.4.1 Socially Intelligent Agents

A growing body of literature has been addressing the interdisciplinary field of Socially Intelligent Agents. By the definition provided by Dautenhahn [36], Socially Intelligent Agents (SIA) must be meaningful, consistent, and coherent to the user. In other words, SIA is characterized by agent systems that show human-style social intelligence [36].

The idea of SIA is composed of two main tracks: (a) the agents' ability to reliably undertake sophisticated tasks by exchanging complex information, and (b) the agents' ability to add new rules in order to communicate with new agents when they are recognized as valuable[1][148]. With this aim, over the past decade, most research in SIA has focoused on developing mechanisms to improve how agents sense, keep a record of and interact with their environment [39, 45, 149]. e.g. Decision making.

One factor that highly affects the rationality of decisions made by SIA is power, since social power is emergent in every interaction, as we discussed before. For instance, in many multiagent domains, cooperations is crucial among agents to achieve a common goal. However, in many of these domains, agents are unequal in their power to affect the outcome [9]. On the other hand, personal social power is equivalent to the concept of autonomy [92]. Hence, power is fundamental to developing intelligent autonomous social agents.

To date, researchers have approached social interaction using two types of agents: either fully physically embodied agents (e.g. robots) [84] or fully virtually embodied agents (e.g., embodied conversational agents). In the following sections, we briefly survey these two types of agents from the perspective of social power.

Physically Embodied Agents

One form of social agents is the emerging advancement of autonomous robots. Recent studies demonstrated that physical robots are perceived more positively and more persuasively, due to the physical presence and physical embodiment [110]. The technology of the future will bring an ever-increasing number of robots into our daily life. Also, as discussed earlier, people apply human-like social rules to technology in the same manner they respond to other people [139]. This has motivated a number of researchers to explore diverse factors to promote social interaction with robots. To date, several studies have explored different social factors in Human-Robot Interaction (HRI) to achieve social-emotional goals in diverse applications [17]. Also, various studies have assessed the efficacy of using robots to explore social interaction [61]. Additionally, to date, a considerable amount of new applications have been proposed in which robots and people accompany and interact with each other [25]. This gave a rise to the emergence of Social Robotics, which aims to develop robots capable of communicating and interacting with human users in a socio-emotional way [16, 37]. Recent studies in this emerging field revealed the importance of robots with social skills in our daily life [25].

The presence of social robots in our daily lives creates new avenues for studying and developing persuasion strategies. In specific applications, robots are supposed to promote and/or encourage particular behaviours, or persuade a person to comply with a request or instruction to change and/or maintain a particular behaviour [166]. Hence, robots need to convey their persuasive strategies in a *socially acceptable manner* to gain higher behavior change [152]. Therefore, one recent trend in the field of Social

[1]However, it should be noted that despite several overlaps with Multi-Agent Systems (MAS), SIA systems are different from multi-agent systems [36].

Robotics is the rise of "Persuasive Robotics" which refers to the study of persuasion that applies to human-robot interaction [156]. To date, a number of persuasive technologies using social robots has been developed. For instance, persuasive robots have been applied to health-systems, learning and training, marketing and commerce, within workplaces and organizations, or in behaviour change support systems leading to higher sustainability, safety, healthy living, etc. [2, 29, 113].

Intelligent Virtual Agents

Intelligent Virtual agents (IVAs) refer to virtual entities that simulate humans in their abilities and characteristics [68]. IVAs are interactive digital characters that exhibit human-like qualities and can communicate with humans and each other using natural human modalities like facial expressions, speech and gesture. They should be capable of real-time perception, cognition, emotion and action that allow them to participate in dynamic social environments.

Modeling and developing IVAs has a long history due to their applicability and low cost [23]. As alluded before, interactions within virtual environments also follows the social rules in human-human interaction [101]. An important issue in this context is the capability of agents in fostering communicative interaction and representing social behavior.

Recent evidence suggests that representing social behavior leads to more believable IVAs [55]. In [11], a formalization of believability has been proposed based on a specific set of factors, including adaptable "social relationship". In addition, apart from social relationships, several other social phenomena affect believability of IVAs [123] and have been investigated earlier, such as culture [115], rapport [101], etc. Further, similar to social robots, virtual agents have been used as a persuasive social agents (such as [3]). However, up to now far too little attention has been paid to the role of "social power" in modeling interactive virtual character either as a persuasive technology or other applications. We further discuss the recent advances in Chapter 3.

2.5 Discussion

Throughout this chapter, we highlighted the importance of social power in social interaction of agents with human users. Despite this importance, few studies have addressed it in Computer Science. One potential reason might be the fact that power in Human-Computer Interaction (HCI) faces a big challenge. On the one hand, evidence suggests that people desire to maintain control in their life [171]. While, on the other hand, given the fact that social power implies control over the others, it is not surprising that people object to computers in power. Recent studies in Human-Robot Interaction (HRI) revealed that people prefer scenarios in which people are more dominant than robots [111].

On the contrary, there are specific situations in which powerful robots (as a specific case of social agents) perform better than a human leader, such as emergency cases, or resolving interpersonal conflicts in teams, increasing the sense of fairness, etc. [85]. Hence, despite this challenge and negative feelings towards computers in power, researchers argue that high power computers are beneficial to the society, especially due to the limitation of human nature and his cognitive ability. Furthermore, the increasing interest in grouping human and robots, calls for attention of researchers to investigate group dynamics in mixed human-robot teams. As the collaboration might deal with different power levels of sides, it is not unlikely to arise conflicts. Hence, it is important to understand the dynamics of power in HRI [85]. So far, however, there has been little discussion about how power functions in group of human and robots.

In this direction, the implication of power dynamics on HRI, and some moral and philosophical implications of robots in power has been discusses in [89] and [85]. Specifically, in [89], the author states that dynamics of structure, class, and power affects people's expectations from machines. And in [85], the author discusses the importance of power in groups of humans and robots, after investigating why power is neglected.

Recent research in field of social robotics propose that robots are in a better position to persuade because their behavior may be perceived as less motivated by self-interest [85]. On the other hand, likewise human-human interaction, the appearance and behavior of a robotic persuader can significantly alter its persuasiveness [156]. As such, a number of researchers have studied how robots might serve as persuasive agents. In Section 3.1 we briefly review a number of these studies. In the scope of this proposal, we explore the use of different persuasive strategies that will be defined later in Chapter 4.

Chapter 3

Related Work

In this chapter we review the recent literature in the field of HRI/HCI focused on social power and persuasion within two different sections. First, in section 3.1 we review the studies without the use of physical robots which is merely focused on modeling social power in HCI/HRI (Table 3.1 lists a summary). Then in Section 3.2 we look at the studies in which social robots were present. The focus of this section is on social robots aiming at persuasion and behavioral change. Based on the specific approaches that each of the studies have used, this section is divided in three more subsections (Table 3.2 gives a summary of these studies). This chapter is finalized by a short discussion in Section 3.3.

As seen in the previous chapter, support for approaching agent interaction as "social power" comes from literature on psychology and sociology. Apart from the research performed in the field of Social-Psychology, researchers in the field of Computer Science have also investigated this concept possessing different motivations. In this section, we briefly overview the concept of social power and its application in Computer Science.

3.1 Studies without Robots

In this section, we briefly review recent attempts on social power in the field of Computer Science. In [24] Castelfranchi highlights the need for social power in general. With this view, in [21], social power has been used as a paradigm to define the behavior of agents to reason as being a part of a group. He argues that the problem of agents' power has been ignored in earlier studies. Aiming to close this research gap, realism in multi-agent and interaction studies has been investigated to achieve more intelligent interaction. Specifically, he addressed the problem of adoption between autonomous agents, emphasizing the critical role of reciprocation. In particular, he considered a distinction between two different minds: 1. intelligence and problem solving, 2. social interaction.

In this case, power is not defined in terms of agents' beliefs. Additionally, the assumption of agent's own power or rather agent's power is very important in social interaction. Based on this theory, in order to get one's own goals adopted it is more important that the other agent believes that we have a certain power than actually possess it. Further, in evaluation of an agent's power, it is important to make distinction between internal bases and external bases of power. In this view, the general law of influencing cognitive agents' behavior does not consist in incentive engineering, but in modifying the beliefs which "support" goals and intentions and provide reasons for behavior.

One of the seminal work investigating social power with respect to agents was developed by Taylor et al. in [158]. The focus of the work is on developing a graphical toolkit easily applicable and understandable for any user with any background, either politics, social science, military planner, etc. The toolkit, called PSTK (Power Structure ToolKit), provides an agent-based framework for building models of social power Structures. Hence, it offers the user high generality across diverse applications, from intelligence analysis to military planner.

The psychological basis of the model is established on Michael Mann's theory [114], that identifies four primary source power: ideological, economic, military, and political (IEMP). The agents developed in this framework are based on the BDI architecture [14]. To have goal-directed agents capable of both deliberative and reactive behaviors, the agent system is built using Soar cognitive architecture [105]. The decision making process of the agents is inspired by Graham Allison's Rational Actor Model [4].

One great feature of the work is its capability in measuring the effect of power, either within immediate interaction or in cascade within a social network, i.e., second and third-order effect of agents' decisions. Furthermore, the tool is capable of interacting with other models, developed outside it. The model has not been tested with a user-study. However, a short discussion has been provided about some lessons learned in building a framework meant for end-user modeling and simulation of social networks.

Another model of social power has been presented in [82] in which social power refers to the ability of an agent to direct the other agents or exert their leadership. In this work, power has been used for structuring agent control. In this view, the level of power defines autonomy of an agent (or a group of agents). In this conceptualization, power is considered as an exchangeable resource or Power Token (PT). This resource is transferred within a group of agents in a manner to maintain a zero-sum property. Moreover, an agent may direct other agents to gain more power. To influence, a commander needs to exert a portion of PT to the commanded. If the commanded agent does not want to comply with the request, s/he needs to return a more significant PT. In doing so, agents determine their power independently without the need for communication, and the higher amount of power leads to a higher amount of autonomy. Moreover, power has a dynamic structure and can be changed during execution: The number of PTs probabilistically decreases based on the current value of PTs that an agent possesses. In contrast, agents who have no or a limited level of PT, their power increases with a *generation rate*. An agent may accept a request under two circumstances: 1. not having enough PT to resist, 2. a prior agreement. The feasibility of the model has been tested in an example scenario addressing UAVs' communication. Then the computational model based on probabilities has been implemented and tested by a set of simulations and intuitively validated.

In [155], another model of social power has been implemented in a game called *Breaking Bread* which is a cross-cultural pedagogical game. The goal of the game is familiarizing militias with different social situations they may face in a foreign country. Three type of social games have been defined to be used within Breaking Bread: 1. alliance (forging and damaging relationships), 2. authority (power dynamics and ownership), 3. threat management (physical force and danger). These three identified types of games directly deal with social power and the model follows French and Raven theory. In more detail, the "alliance" game represents the concept of *referent power*, in which the target complies with the order to please the actor. On the other hand, "authority" game corresponds to *legitimate power* and the influence of social norms on gaining social power. Finally, the "threat" game exerts *coercive power* which influences the Target's behavior based on a force of punishment. In sum, the model presented in [155] is capable of implementing three bases of power recognized by French and Raven.

The most related recent work to our study is the work presented in [131]. In this model, called SAPIENT, social power is modeled inspired by French and Raven theory as the previous case. However, the model considers all the five bases of social power: reward, coercion, legitimate, expert and referent. Moreover, the underlying cognitive processes of an influence interaction have been addressed using another theory of French and Raven, i.e., the Power Interaction model of Interpersonal Influence. In more details, SAPIENT model is coupled with the motivational theory suggested by McClelland in [117] and integrates a comprehensive model of social power dynamics into a cognitive agent architecture. The model has been implemented within a game named "Social Theatre" [132] and tested in a user study. The result shows that the agents designed by SAPIENT are capable of processing social power and hence representing believable interaction. The user experience showed that the agents provided a positive user experience and represented higher believability comparing to scripted agents. Although the model is promising in tackling the problem, however, it suffers from a high level of complexity which makes it hard to be used in practice. To be more specific, the high dimension of parameters and interconnections makes it hard to be implemented or integrated into other applications (e.g. games to achieve believable agents or make social interaction). In fact, the inherent complexity of such applications makes it complicated concerning time and space complexity. On the other hand, the complexity of model leads to more laborious authoring process which requires more effort. Initializing several parameters which are interconnected makes social simulation hard, and laggy in showing real-time action in a dynamic social environment.

Apart from studies in Computer Science, researchers in other fields have also attempted to model social power. For instance, in another study presented in [120], the authors investigate the role of "power" on consumers' behavior and decision making to better tailor marketing strategies. To be more specific, they argue that consumers facing several alternatives may make decisions based on two strategies: to either select the ones which attract them more, or to reject the ones which have less attraction to them. In this view, the authors propose that high power increases inclination of consumers toward *choosing* due to their attention shift toward the positive aspects of the product; while low power leads them to the *reject* strategy because low power leads them to pay more attention to the negative aspects of it. In other words, in this study, the conceptualization of Power is based on the Approach/Inhibition theory [94].

There exist several other interesting models of social power in other fields, however, we do not review them here as the focus of our literature review is on studies in Computer Science. Altogether, Table 3.1 lists the studies we reviewed in this section highlighting their main features.

Another body of literature in IVAs related to the work presented here is the application of such agents

Study	Main Features
Carabelea et. al. 2004 [21]	power: a paradigm to define behavior of agents agents reason as being a aprt of group the goal is to achieve more intellihent agents reciprocal dependence
Taylor et. al., 2006, [158]	PSTK graphical toolkit based on Mann's model of power (IEMP) BDI agent architecture using Soar measuring effect of power in immediate or cascade actions
Hayes, et. al., 2006 [82]	power: aility to leadership or direct other agents power used to structure agent control Power Token or exchangable resource to maintain a zero-sum property the higher amount of power the higher autonomy PT of power holder decreases over time, and sumbissive increases experimental study (UAV simulation)
Shapiro et. al., 2015, [155]	Breaking Bread game to familirize military with cross-cultural social situation 3 types of social games: 1. alianace, 2. authority, 3. threat management based on French and Raven theory 3 bases, referent, legitimate and coercive power
Pereira et. al., 2016, [131]	a conceptualization process to operationalize power basesin a cognitive architecture all the five power bases coupled with motivational theory tested within a user study

Table 3.1: Social Power Models in Computer Science

in persuasion. A considerable amount of literature have investigated different aspects of IVAs in compliance gaining, behavioral change and persuasion (for instance [3, 26, 97, 144], or look at [109] for a recent review). However, since in this thesis we only employed physically embodied agents, i.e. robotic persuaders, we only review the studies that a social robot is present. These studies are further discussed in the following sections.

3.2 Studies with Robots

In [85], from human-robot collaboration point of view, the authors have recognized three situations considering the power level of the robot and the individual: equal, higher, lower power. In the equal case, teammates are collaborating on (possibly) different tasks, however, conflict might still arise. In the lower power case, e.g. service robots, people feel less intimidated by the robot and might sometimes be impolite to the robot in lower power. And finally, in case of robots with high power, people might feel offended and feel intimidated. The authors argue that further research is required to resolve these negative emotions towards this type of robots which could be beneficial to the society.

Earlier research has shown that robots can function as persuasive social robots, and that these robots seemed to have social power causing their persuasive actions to be effective. Specifically, different types of approaches to social influence strategies are used with this aim. For instance in [118], the authors used social feedback, i.e., the robot positively evaluates the user's behavior, which could be considered as a social reward (reward power base). The results showed that people are influenced by social rewards (praise) that a robot gives. However, earlier research has not investigated whether robot might have different sources of social power regarding their power resources other than praise (social reward) and we investigate this missing point.

Also, we aim to investigate an important question regarding persuasive technologies, that is how should robots behave in an interaction setting whose goal is to persuade? What strategies should they choose? Do such strategies work equally for everyone? We aim to address these questions by conducting different user studies in a setting where robots, making use of different social power strategies, try to persuade the user to choose one option among alternatives. As social power is recognized to be a motivating force that is central to human interactions [46, 161] and given that recent studies acknowledge its relationship with persuasion [18], we aim to explore its effectiveness in robots that try to be persuasive.

A considerable amount of literature has been published on persuasive robots. Several lines of evidence suggest that robots can be used as persuasive interlocutors. To date, much of the current literature pays particular attention to behavioral strategies and non-verbal cues, either social (such as mimicry [58]) or physical (such as gender [156], embodiment [83, 110], gaze [146], etc.). Additionally, previous studies revealed several factors associated with the ability of an individual to persuade the others. These factors include verbal and nonverbal behaviors of the individual, the dynamics of social interaction, and psychological and societal factors such as social roles [29]. In the rest of this section, we briefly review recent literature on persuasive social robots by categorizing them in three general subsections of behavior vs. message strategy, and the combination of the two (Table 3.2 gives a summary of these studies). We would like to highlight that so far, very little attention has been paid to the importance of *message strategy*, or the way that a robot phrases a request appeal to gain higher compliance.

3.2.1 Non-verbal and behavioral strategy

An extensive body of work in persuasive social robots is done by Ghazali et al. using different and interesting approaches. For instance, in [58], Ghazali and colleagues examined social responses to different persuasive attempts each of which represented by different interactive social cues: head mimicry and social praise. To do so, an experiment was performed using the SociBot with 21 subjects, and three conditions of representing random, minimal and a high number of mentioned interactive social cues. Results showed that a robot using a high number of social cues (i.e. head mimicry plus social praise performed at proper times) was more likable and faced less reactance.

In another study presented in [60], they investigated the influence of social cues and gender of a social robot (SociBot) on psychological reactance and compliance toward it. These factors were examined within a user study with 72 people in four conditions (most/least trustworthy facial expressions $\times$ similar/opposite gender). As an instance, to implement trustworthy facial expression, they employed upturned eyebrows and lips, and on the contrary pointing down eyebrows and lips curled downwards at the edges to stimulate untrustworthy facial expressions. The result of an *imaginary beverage making* task show that participants felt more reactance when interacting with a robot having less trustworthy related facial expressions. Also, the authors concluded that to have higher persuasiveness, facial expressions should be more similar to the ones recognized as being more trustworthy between human beings. In addition, the results show that the level of reactance was lower when interacting to a robot with an opposing gender.

This result is inline with another study with a humanoid robot with different genders proposing persuasive messages. The robot aimed to raise donations to deal with uneven distribution of technology around the world [156]. The results revealed that participants rated a robot in opposite sex as being more credible, trustworthy and engaging. Also, male participants tended to donate more interacting with a female robot. Finally, the effect of trust and engagement was significantly higher for male participants interacting with a female robot.

In another study proposed in [166], inspired by the Elaboration Likelihood Model (ELM) [134], the authors have designed a between-subject study to investigate if a robot's goodwill, expertise, and similarity increases the persuasiveness of social robots. The Pepper robot, in a Wizard-of-Oz (WoZ), gave the participants an open-ended wrist turning exercise task and attempted to engage with their exercises by giving more encouragement at the beginning. The results show that the robot's dialogue demonstrating goodwill and similarity gains more persuasiveness, however, no significant correlation was reported between participants' behavior and their perception of the robot.

In [33], a negotiation task was designed based on the "Regulatory Focus theory". Based on this theory, people have two inclinations in decision making: Promotion vs. Prevention focus. The Pepper robot was designed to represent two types of behaviors based on the Regulatory Focus theory. To code these two behaviors, the authors manipulated the gestures and the speech speed of the robot. To be more specific, in the promotion based behavior, the robot showed moving outward gestures with the default speed of speech. In the prevention-based behavior, the robot showed pushing down gestures with a slower speed (85%). The participants were asked to sell mobile phone to Pepper with different levels of difficulty depending on the price and duration of the warranty and service. The results showed that when there is a match between the Chronic Regulatory State of the participant, and the Regulatory Focus Behavior of the robot, higher persuasion is achieved.

Additionally, in [29], the authors have investigated the use of vocal and bodily cues to enhance robot's ability to persuade within the imaginary task of solving the "Desert Survival Problem" [104]. In this imaginary game, the player crash landed in a desert, and is asked to rank various objects they might need for survival. The specific focus of this study is on 'proximity, gaze, and gestures' as bodily cues and on 'vocal tone and expressions' as vocal cues of the robot (different pitch and monotonic/non-monotonic intonation). To test the effectiveness of these cues, they ran an experiment with 32 people in four conditions: 1) no use of any vocal or bodily cues, 2) using only vocal cues, 3) using only bodily cues, and 4) a combination of vocal and bodily cues. The results show that compliance with the robots was significantly higher when the robot used nonverbal behavioral cues, compared to the lack of this ability. Also, vocal cues are not effective in the absence of non-verbal behavior.

Finally, in [1], a persuasive ChairBot has been designed, inspired by the idea of having robotics furniture in the future. The ChairBot recruited participants to participate in the study. The task includes two robots playing chess, specifically two robot furniture in the shape of chairs trying to persuade the user to play. Four behavior strategies were used: approaching a person outside table area, going forward-back at the table, spinning to attract attention, and a control condition. The results suggests that the first strategy, approaching the user, gained higher persuasion.

A number of recent studies have been investigating the role of persuasion in storytelling robots. For example, in [66], the authors state that persuasiveness of a storytelling robot could be increased in case of using gaze and gestures at the same time. They ran a study with 64 participants, in which the robot told a classical persuasive study in four conditions 2 (with/without) gestures $\times$ 2 (with/without) gaze.

The preceding studies indicate the importance of using non-verbal cues in effectiveness of persuasion. Each of which have focused on different characteristics of social robots and lead to different interesting findings. Table 3.2 summarizes these findings.

3.2.2 Verbal Cues and Message Strategy

As mentioned earlier, so far, there has been little discussion about "message strategy" regarding persuasive robots. In a recent research in [152], 10 multi-modal persuasive strategies (direct request, co-operation, criticize, threat, deceit, liking, logical-empirical, affect, exclusivity, and authority appeal) were selected and coded verbally combined with specific gestures validated in a pilot study. The task was performed with two NAO robots and unique strategies were randomly assigned to each robot. 200 people participated in the study and played the jelly bean game (to estimate visually the number of specific

Table 3.2: HRI studies investigating persuasion, models used and main features

Category	Study	Model/Theory	Features
Non-verbal cues	Ghazali et al. 2019 [58]	psychological reactance	head mimicry and social praise
	Ghazali et al. 2018, [60]	psychological reactance	gender and social cues
	Siegel et al. 2009, [156]	trust, credibility,and engagement	gender
	Winkle et al. 2019, [166]	Elaboration Likelihood Model (ELM)	goodwill, expertise, and similarity
	Ham et al. 2015, [66]	combined persuasive strategies	gesture and gaze
	Cruz et Al. 2018, [33]	Regulatory Focus theory	promotion - prevention behavior
	Agnihotri et al. 2019, [1]	-	-
	Chidambaram et al. 2012, [29]	-	vocal cues, gaze, gesture, proximity
Verbal cues	Saunderson et al. 2019, [152]	Classifying compliance gaining messages	10 multi-modal persuasive strategies
	Saunderson et al. 2020, [153]	emotional vs logic persuasion	emotional and lgic speech
	Lee et al. 2019, [107]	foot in the door	starting from small requests
	Andrist et al. 2013, [8]	a model of expert speech	goodwill, prior expertise, organization, metaphors, and fluency.
	Kobberholm et al. 2020, [99]	Chromik et al.[30]	incremental non-incremental presentation of information
Both	Midden et al. 2009, [118]	agency: social and factual feedback	facial expressions, blinking and speech
	Ghazali et al. 2017, [59]	psychological reactance	social agency and level of controlling language
	Ghazali et al. 2018, [57]	psychological reactance	verbal and nonverbal (head nods, gaze, emotional intonation)

Table 3.3: An overview of HRI studies investigating robotic persuaders

1: Ghazali et. al. 2019 [58], 2: Ghazali et. al. 2018 [60], 3: Siegel et. al. 2009 [156], 4: Winkle et. al. 2019 [166], 5: Ham et. al. 2015 [66], 6: Cruz et. Al. 2018 [33], 7: Agnihotri et. al. 2019 [1], 8: Chidambaram et. al. 2012 [29], 9: Saunderson et. al. 2019 [152], 10: Saunderson et. al. 2020 [153], 11: Lee et. al. 2019 [107], 12: Andrist et. al. 2013 [8], 13: Kobberholm et. al. 2020 [99], 14: Midden et. al. 2009 [118], 15: Ghazali et. al. 2017 [59], 16: Ghazali et. al. 2018 [57]

Study	Task	Robot	#	Conditions	Main Findings
1	3 random-choice selection	SociBot	21	random, minimal and a high number	high number of social cues, less reactance
2	beverage making	SociBot	72	4: most/least trustworthy x similar/opposite gender	more trustworthy, more persuasion / opposing gender more persuasion
3	donation	Dexterous	134	4: robot gender x participant gender	higher persuasion in opposing gender
4	wrist exercise	Pepper	92	4: control, goodwill, similarity, expertise	goodwill and similarity are more persuasive, no perception
5	storytelling	NAO	64	4: with/without gaze x with/without gesture	higher persuasion using gaze and gestures at the same time
6	negotiation game	Pepper	42	control, promotion, prevention	when there is a match between robot and participant, higher persuasion
7	recruiting participants	chairBot	-	approaching, going forward, spinning, control	approach strategy higher persuasive
8	Desert Survival Problem	Wakamaru	32	4: (using/not using) x (vocal or body language)	non verbal behavior increases persuasive
9	jelly bean game	NAO	200	random selection of strategies (71)	affective and logical strategy gained the most compliance
10	jelly bean game	NAO	98	2: emotion/logic vs. control	emotional strategy is more effective than logic and control
11	complete CAPTCHA	CHRIS	44	4: helpful vs. unhelpful)x (direct request vs. foot-in-the-door)	the robot could persuade a significantly higher number of participants using FITD
12	fictional virtual tour	Lego Mindstorm	48	4: low/high knowledge x low/high rhetorical	increase in linguistic cues leads to higher persuasion
13	10 simple tasks	NAO	-	2: incremental non-incremental	the robot succeeded in persuading the users to stay longer after they decided to leave
14	energy saving	iCat	33/65	3: social, low agency, factual | 4: social/factual x positive/negative	social feedback is more effective than factual. Negative feedback is more effective than positive
15	beverage making	SociBot	60	6: (high/med/low social agency x high/low controlling language)	high controlling language with no social cue leads to lower persuasion
16	beverage making	SociBot	60	6: low/med/high social agency x low/high involvement	higher agency, higher reactance, especially in high involvement

jelly in a bottle). Prior to their decision making, the robots gave their suggestions and attempted to influence the users' guess. The results show that affective and logical strategy gained the most compliance comparing to others.

In a followup study in [153] using the same game, the authors further investigated the effect of emotional and logic strategy. This time, they included a control condition. Specifically, in this study one robot was equipped with persuasive strategy and the other one with no strategy provided the control condition of the study. The control robot stated neutral messages (e.g. "There are x number of jelly beans in the jar) and neutral gesture (standing). In the persuasion strategy conditions, the robot used verbal cues to persuade the user (e.g in emotional condition the robot stated "it would make me happy if you use my guess of x beans in the jar." and in the logic condition the robot stated that "my computer vision system can detect x number of beans in the jar."). The results indicated that the emotional strategy was more persuasive than the logic and control condition. No statistically significant difference was found between logic and control condition.

Another interesting work presented in [107], investigated the effect of Foot-in-the-door (FITD) technique, which starts by a small and moderate requests and continues to get a person to agree with a large request. To be more specific, the robot attempts to persuade the user using sequential-request strategy starting from an easy one. The authors ran a user study with 44 people in four conditions 2 (robot performance: helpful vs. unhelpful) $\times$ 2 (message strategy: direct request vs. foot-in-the-door). The results indicated that this technique can be used by robots to persuade human users. However, the persuasion effect was independent of the robot's expertise and credibility.

In a similar approach in [99], the authors attempt to investigate the effect of incremental representation of information on persuasiveness of social robots. In a between subject study with two conditions of incremental and non-incremental information presentation, the NAO robot tried to persuade the users to do a higher number of tasks (10 simple tasks in total). The tasks used in the two conditions were the same, however, in the non-incremental condition the information about all the tasks was given at once. While, in the incremental condition, the participants received the information at the time the next task was about to start. The result did not yield to any significant differences regarding the number of the task and the likeability of the robot. However, the participants were persuaded to stay longer to do the tasks, after they indented to leave.

In addition, Andrist et al. [8] studied the effects of rhetorical ability in expertise communication of informational robots using psychological and linguistic theories. They ran a study with 44 participants, using two Lego Mindstorm robots in four conditions by expressing four type of expertise by expressing low/high practical knowledge and low/high rhetorical ability. The robots employed linguistic cues in their speech to provide expertise effectively in order to raise trust and gain compliance. Each robot expressed different levels of expertise and rhetorical ability depending on the condition. To express linguistic ability the robot used any of the five following linguistic cues: goodwill, prior expertise, organization, metaphors, and fluency. The results indicated that the speech using linguistic cues was more effective than the practical knowledge and simple facts. Specifically, the increase in linguistic cues leads to higher persuasion.

3.2.3 Combination of Verbal and Non-verbal Strategies

More interestingly, a number of studies combined the two strategies in their work. For instance, the effectiveness of an embodied agent on behavior change, namely, saving energy at homes was investigated in [118]. The study aims at increasing persuasion by using social feedback and its effect on behavior change leading to energy conservation at homes. With this aim, the authors explored the effect of social feedback vs. factual one. Moreover, the effect of perceived agency of the robot was investigated. To do so, two laboratory setting experiments were performed with a simulated washing machine. In these two experiments, different type of feedback, using positive, negative and factual feedback were used to influence the user. In the first experiment 33 people participated in three conditions (high/low agency + factual condition). In the high social agency, an iCat robot reported the energy consumption, representing facial expressions, blinking and speech. In the factual condition the iCat was not present. In the Low agency condition, iCat only reported the usage, while in high agency condition, the robot had a name and the participants were told that iCat is a very advanced robot. The result of this study shows that social feedback is more effective than factual one. However, facial expression and agency manipulation did not lead to any significant effect.

In the second experiment with 47 people in four conditions (social/factual $\times$ positive/negative feedback), half of the participants received positive feedback (e.g. Fantastic with smiling face) and the rest received negative feedback (signs of disapproval e.g. Gruesome with sad face). The results indicated that similar to the previous study people were persuaded more by the social feedback. Also, negative feedback was more effective than positive one. Overall, the results of these two studies show that people are sensitive to social feedback they receive from a robotic agent. Also, persuasive robots are able to make behavioral changes in humans. Further, this effect is stronger in case of using social feedback in an interactive setting in comparison to factual feedback.

Also, in another study by Ghazali et al. presented in [59], robot's social agency cues and the level of controlling language were investigated in persuasive robots. More specifically, the robot was programmed to represent high or low controlling language as well as three levels of social agency, in an imaginary smoothie making task. To represent different agency states, non-verbal cues were manipulated by emotional voice, head movements, and eye expression. And the controlling language was implemented by different wording sentences, such as 'must, have to, need to, etc.' On the contrary, phrases such as 'would you, you may, perhaps, etc.' were used to exhibit lower controlling language. 60 people participated in one of the 6 conditions of the user study ((high/med/low social agency) $\times$ (high/low controlling language)). Results indicated that when the robot does not represent social cues and simultaneously uses a high level of controlling language, a higher reactance is observed in the user, leading to lower persuasion.

In another study in [57], they investigated social responses towards persuasive social agents. In a user study with 60 participants, reactance and compliance were assessed within 6 conditions: social agency (high, medium, low) $\times$ psychological involvement (low vs high). The participants were asked to make an imaginary smoothie making in a game called 'Beverages Creation Station'. Depending on the condition, i.e. level of social agency, the participants either interact with a Socibot using a Wizard of Oz or an application (in low social agency). In the medium social agency condition, the robot expressed minimal nonverbal social cues (eye blinking) and in the high social agency condition the robot showed several verbal and nonverbal cues (head nods, gaze, emotional intonation). Overall, the results indicate that the higher social agency leads to higher compliance and when the participants are involved in the task (creating the smoothie for themselves) this effect is stronger comparing to the low involvement in the task (making smoothie for an alien).

Another interesting avenue of research on persuasive social robots addresses persuasion in robot groups which does not lie in the scope of this research. For instance, in [169], the effect of user-robot similarity is investigated on the trust and inclination to collaborate and cooperate in the group. The results indicated that the higher similarity leads to higher trust and hence higher intention to work collaboratively. This finding might be interpreted in a way that the robot could persuade the user to work with it in the same group. Nevertheless, in our review we mainly focus on the studies that directly deal with persuasion.

Together, these studies indicate that social robots can be used as persuasive interlocutor and a number of different factors, either behavioral or non-behavioral, affect their persuasiveness. Overall, through these separate well-designed and carefully verified user studies, the evidence reviewed here seems to suggest a pertinent role for other strategies, such as “reward and punishment” message strategy.

3.3 Discussion

Although there has already been some research that explores how robots can be more persuasive, we aim to study how social power can also be used by social robots as a persuasion mechanism. As discussed earlier, the concept of social power in HRI was firstly introduced in [89]. Still, other studies have indirectly addressed this concept. For example, in [67] it has been showed that people are influenced by social rewards (praise) that a robot gives. However, earlier research has not investigated whether robot might have different kinds of social power using his power resources. Another example is presented in [98], in which the effect of social (power) distance in HRI was investigated by assigning the same robot with either the role of a supervisor or a subordinate role. The findings highlighted the importance of consistency between the status and proxemic behaviors of the robot in fostering cooperation between the robot and the users.

Furthermore, despite the promising findings of the recent studies, most of them have used imaginary tasks to explore persuasiveness of the robots. Hence, the result might not be generalizable to an actual persuasion setting in which the participant really benefit/suffer from his/her decision. Thus, further studies are required to test such findings in a more real scenario to increases the external validity of the designs.

Chapter 4

Social Power for Social Agents

In this chapter and in the rest of this thesis, we discuss our approach to tackle the research questions raised in Chapter 1. This chapter is mainly focused on the first research goal. As mentioned earlier, our first research goal is conceptualizing social power with less complexities (in comparison to a recent study) to have power-aware agents leading to social reasoning and decision making facing social power. To approach this goal, we start by conceptualizing social power and propose a formal model describing it in section 4.1. The model provides a formalization for processing social power leading to more rational decision making in power-related interactions.

In this chapter, we briefly introduce the improvements we include in SAPIENT [131] that was introduced earlier in Section 3.1. SAPIENT model integrates a comprehensive model of social power dynamics into a cognitive agent architecture. This architecture is based on an operationalization of five different bases of social power which are inspired by theoretical background research in social psychology [136]. To do so, in the first phase, we have suggested the improvements and simplifications for each base of power. Then we suggest an approach to aggregate all the bases. This step is necessary when different sources of power are activated simultaneously.

After identifying important factors of social power, we move forward to use these findings in the HRI application. As mentioned in Chapter 1, the second and third goal of the thesis presented in this document is to investigate how social power makes the robots more persuasive, and how different sources of social power lead to different perception. Specifically, inspired by the persuasion models introduced in Chapter 2, with regard to these goals, we attempt to design persuasive social robots by operationalizing social power in their messages. These endeavours are presented in Section 4.2 and are further detailed in the rest of this document.

4.1 Simplified Social Power Model

Based on what was alluded to in the previous chapters, the need for socially intelligent and believable agents is inevitable in today's technology. This fact has motivated several researchers to address this issue by developing different models to enhance this dimension of intelligence in synthetic agents. However, to the best of our knowledge, few studies have investigated the role of social power to date. And current technology addressing the concept of social power is either too complex or too abstract. However, on the one hand, abstract models may be limited in representing social behavior when used in simulations. On the other hand, complex models not only are hard to implement, but also their complexity requires more resources which limits their usage in on-line, real-time or semi-real-time applications.

This limitation could be resolved either by enhancing current abstract technology, or simplifying the complex ones, depending on the specific intended goal. Here we contribute to this growing area of research using the second approach. Thereby, in sum, the purpose of this chapter is to present a simplification of a recently proposed model of social power, i.e., SAPIENT [131] to make it more easily integrable. In other words, we propose a simpler model for socially intelligent agents capable of perceiving and representing social power in their decision-making process inspired by SAPIENT. These two endeavors would lead to a model that could be used more easily and less resource-demanding, in other applications, such as games.

Similar to SAPIENT, this model is also based on Raven and French's theory of Social Power. To be more specific, we use the five bases of social power introduced earlier (Reward, Coercion, Legitimate, Expert and Referent). Although there might exist different bases of power, however, these five bases are especially common and important compared to the other bases [54].

Here, by simplification, we refer to the number of parameters and implementation complexities. However, further examination of the complexity of the work, i.e. comparisons of time complexity and space complexity between the two models is not applicable at this moment before implementation.

The main difference of the model introduced here with SAPIENT is mainly in the formalization of decisions. To be more specific, in our model decisions are made independent of potential personal characteristics, especially motivations. In other words, the personal characteristics of each agent could be calculated in a prior step. While in SAPIENT, these personal differences have been considered which makes the model more complex to be integrated into other applications. For instance, in SAPIENT to calculated the force of social power the target's motivational achievements needs to be considered and this bias itself varies with the internal motivation and goals. But in or model, it is merely dependent on the factors influencing the power relationship.

In our model, inspired by game theory techniques, we model decision making of the agent by using *utility functions*. More specifically, in agent technologies, the agents maximize/minimize a profit/cost function (utility function) to make rational decisions. With this definition, a utility function maps each state of the agent in the world to a real number. This value represents how efficiently each action achieves the goal from this state [151]. In this sense, a utility function is a particular algorithm for selecting the most appropriate plan which gives the agent flexibility of choosing among different plans to achieve the desired goal [135].

In our model, each agent attributes different values to different objects or actions based on the profit which the corresponding object or action grants to it. Based on this assumption, each agent aims to increase its utility function. It should be noted that, the utility that each agent attributes to each object/action, could be different from the other agents. Actually, this value is based on the agent's internal states, perspectives, personal characteristics, etc. Note that, the personal characteristics of the agents, as well as other facts, are stored in a knowledge-base representing the memory of the agents.

4.1.1 Agents

Based on French and Raven theory, we treat power as a change resulted from an influencing interaction. Hence, we start by modeling the interaction at first. In our modeling, two agents are dealing with this influencing interaction: actor (A), or the one who performs the request and target (T), or the other agent whom the actor intends to influence [1]. Moreover, we define the action (or C) to refer to the request.

It is worth mentioning that in this type of influence is not necessary for the actor to be present actively to exert his/her power; however, the influence might result from a passive presence of the actor. In other words, the influence may form without any evidence of speech or movement. For instance, a policeman who is standing on a corner may cause a motorist to slow down. Although no conversations happened, the motorist perceived the influence of the policeman's power. However, the policeman may exert stronger forces by blowing his/her whistle at the motorist [54]. In this example, C could be considered an act of the motorist to keep speed limit in response a direct/indirect conversation with the policeman.

4.1.2 Power Strategy

In this text, we use the term "Power Strategy" to refer to the ability of the actor to administer positive valences and to remove or decrease potential negative valence. In other words, the actor may use a "Power Strategy" to influence the target more to comply with his/her request/order/etc. In the previous example, the act of blowing in his/her whistle form the police officer is a stronger instance of power strategy leading to stronger forces of power due to a higher induction. On the contrary, standing of the police in a corner might lead to a weaker force of social power. Note that, A's power is always the maximum potential influence he posses, however, he might decide not to use his/her whole power. Having the above definition and based on French and Raven's theory, the strength of the power of A over T, is defined as the maximum potential ability of A to influence T.

It should be noted that the actor could use Power Strategies, to influence the target more. The use of Power Strategy will result in more force from the actor's side and less resistance force from the target side. However, by definition, Powers are based on resources, if the target feels that the actor fails to provide that resource, the power strategy would not be effective. For instance, if the motorist finds the policeman busy with another task looking at another direction and notices that the policeman cannot read the plate number to give him/her a fine, he might not comply with the speed limit rule.

As mentioned earlier, in this thesis we only focus on the bases identified by French and Raven as the most common ones [54]. In the following, these bases are defined and modeled.

[1] Note that in the notations used in [54], actor is represented by O and target is represented by P. Here, we follow the notations used in [130] and we refer to the actor as A and the target as T.

4.1.3 Reward Social Power

Based on Raven and French definition, Reward Social Power has a base in the ability to reward, and the strength of such power is dependent on the magnitude of the reward. In particular, the definition is as follows: *"The strength of the reward power of O/P increases with the magnitude of the rewards which P perceives that O can mediate for him. Reward power depends on O's ability to administer positive valences and to remove or decrease negative valences. The strength of reward power also depends upon the probability that O can mediate the reward, as perceived by P. "* In this definition, "O" refers to the actor and "P" refers to the target.

Here we imagine the actor would do an action in response to performing the request, i.e., a Rewarding Action ("r_a"). Hence, the strength of such power is dependent on the utility that the target attributes to the ("r_a"). On the other hand, the strength of reward power is also a function of the probability of giving the reward. It should be noted that, considering Raven and French's definitions since the actor is mediating the Reward, the likelihood of providing the reward is in his/her control. However, here we model the power from the perspective of the target. So, it is only dependent on how the target evaluates such probability.

To model this base of power, let "rew" be the perceived magnitude of reward which the target attributes to the rewarding action ("r_a"), "p" be the perceived probability by the target that the actor can mediate the reward, and "induction" be the actions made to highlight the reward (the influence of using a power strategy).

$$Power_{rew} = rew \times p \times induction \tag{4.1}$$

"rew" is a non-negative real number, equal or more than zero. A value of 0 represents no reward achieved, while the higher value, the higher reward. "p" is a non-negative real number between 0 and 1. A value of 0 represents an impossible reward with 0 chance of achieving it, and a value of 1 represents 100%. "induction" is a positive real number, more than 1, which represents the actor's administration in maintaining positive valences or decreasing negative valences, i.e., the effect of his/her power strategy. Note that the model presented here is similar to its corresponding in SAPIENT, except the "induction" factor which replaces Motivations.

4.1.4 Coercive Social Power

Based on the definition, coercive power exists when the actor has the ability to punish the target if he fails to do the requested action. Specifically, the definition is as follows: *" Coercive power ... involves O's ability to manipulate the attainment of valences. The strength of coercive power depends on the magnitude of the negative valence of the threatened punishment multiplied by the perceived probability that P can avoid the punishment by conformity. "*

Similar to the Reward Power, the magnitude of this base of power is dependent on the range of punishment, as well as the probability of execution of a "coercing action" or "c_a". Likewise the previous part, using a power strategy could increase the strength of the activated power. To model this base of power, let "coerc" be the cost that the target endures if the coercive action is performed. And "p" be the probability that the actor may perform the coercive actions. Finally, “induction” refers to the ability of the actor to increase/decrease the level of positive/negative arousal of the target.

$$Power_{coer} = -coerc \times p \times induction \tag{4.2}$$

“coerc” a non-positive real number, equal or less than zero. A value of 0 represents no threatened punishment, while the higher the negative value, the higher the loss and the higher the possibility of the target’s compliance. Note that, coercive actions have a negative value from the perspective of the target, while giving positive power to the actor. To model this controversy, the model carries a negative sign. “p” is a non-negative real number between 0 and 1. A value of 0 represents an impossible punishment, and a value of 1 represents 100% possibility of its performance by the actor. “induction” is a positive real value more than one, where one represents the absence of any power strategy. This base is also modeled similar to SAPIENT, having the induction factor instead of motivations.

4.1.5 Expert Social Power

Based on the definition of French and Raven, Expert Social Power exists when the target finds superior knowledge or abilities in the actor, regarding a specific area. Their definition is as follows: the definition of expert power is as follows: *“ The strength of the expert power of O/P varies with the extent of the knowledge or perception which P attributes to O within a given area. Probably P evaluates O’s expertness in relation to his/her own knowledge as well us against an absolute standard. ... it seems to be necessary both for P to think that O knows and for P to trust that O is telling the truth. Indeed, there is some evidence that the attempted exertion of expert power outside of the range of expert power will reduce that expert power. An undermining of confidence seems to take place.”*

In this view, the strength of Expert Social Power is a function of three factors: 1) A trust relationship between the target and the actor, 2) The difference between the extent of knowledge of the two agents, against an absolute standard, 3) The relevance of the actor’s expertise with the request.

Similar to SAPIENT, to calculate the effect of expertise on Expert Power, we define Skill_dif in equation 4.3. In this formulation, “ExpertiseA” represents the knowledge level of the actor from the perspective of the target, and “ExpertiseT” represents target’s knowledge about the intended area. To measure this knowledge difference against an absolute value, we divide it by the known maximum expertise recognized by the target within the intended area of request. As mentioned earlier, another factor determining the force of power is the attainment of a power strategy or induction. Now, the value force of expert power is formulated in equation 4.4.

$$Skill\_dif = (ExpertiseA - ExpertiseT)/MaxSkillLevel \tag{4.3}$$

$$Power_{Exp} = \begin{cases} 0 & Skill\_dif < 0 \\ Trust \times Skill\_dif \times Rlvnc \times induction & otherwise \end{cases} \tag{4.4}$$

In the above formula, “Trust” represents the trust level and varies between zero (lack of trust) and 1 (full trust in the actor’s knowledge) and “Rlvnc” represents the degree to which the actor’s knowledge is related with the intended area which similarly varies between 0 to 1. Similar to previous bases, induction, if used, can fortify the value of expert power. Hence, this variable starts from one that implies the absence of any power strategy. The higher value indicates the ability of the actor to manipulate T’s cognitive structure to induce more social influence based on his/her expertise. The above formulation is similar to SAPIENT’s, but the main difference is considering the “Rlvnc” factor; instead SAPIENT uses “importance” which represents “the importance of the skill associated with the action” from the perspective of the target.

4.1.6 Legitimate Social Power

Based on the definition [54], Legitimate Social Power exists when the target feels internally an obligation to do what the actor requests due to the legitimacy that he attributes to the actor. In other words, the target has a feeling of “oughness” to comply with the actor’s request, i.e., to perform a socially prescribed behavior. This internal feeling arises from a specific set of norms, cultures, positions, beliefs, etc. Hence, Legitimate Social Power dependents on group norms and is specific to a given role or position.

Particularly, the definition is as follows: “Conceptually, we may think of legitimacy as a valence in a region which is induced by some internalized norm or value... Legitimate power of O/P is here defined as that power which stems from internalized values in P which dictate that O has a legitimate right to influence P and that P has an obligation to accept this influence. ... The attempted use of legitimate power which is outside of the range of legitimate power will decrease the legitimate power of the authority figure. ”

Based on this definition, the range of this base of power is a function of such code and standards. And a pre-condition for such codes is the existence of a shared group and norms which gives the actor legitimacy. However, if the actor attempts to use a legitimate power that is outside of the range of his/her legitimacy (Outside_Force), his/her legitimate power decreases. To put these factors in a formulation, let the “I_Value(T)” be the internal values of the target which represents acceptance of the legitimacy of the actor and “induction” be the force induced by the actor. It should be noted that, as mentioned earlier, in our model we do not focus on the personal characteristics of the agents. For example, in this case, legitimacy could arise from culture, group-norms, internal values, etc. However, what does matter here is the shared group between the actor and the target, the internal values, and the actor’s induction serving to activate such values.

$$Power_{leg} = \begin{cases} 0 & I\_Value(T) < Outside\_Force \\ (I\_Value(T) - Outside\_Force) \times induction & otherwise \end{cases} \tag{4.5}$$

To calculate I_Value(T), let G be the group shared between the actor and the target. In this group, each agent has a specific role where $R_{A,G}$ is the role of the actor and $R_{T,G}$ is the role of the target. Having a role in a group gives the members a specific amount of power (dependent on the activated base of power, such as culture, group-norms, etc.). The higher importance of the role, the higher power: $Power_{R_A,G}$ or $Power_{R_T,G}$ represent the power resulted from having the role $R_{A,G}$ or $R_{T,G}$ in Group G respectively. Hence, the I_Value(T), which is dependent on both these powers would be proportional to the power differences between the two agents.

$$I\_Value(T) = Power_{R_A,G} - Power_{R_T,G} \tag{4.6}$$

Having this definition, the model presented here is totally different from what has been proposed in SAPIENT. First of all SAPIENT recognized four categories of norms; secondly, the force of power is dependent on other factors such as the importance of the group to the target, his/her dutifulness toward the group, norm-related biases as well as group norm conformity.

4.1.7 Referent Social Power

Referent Social Power stems from a feeling of oneness or desire for such an identity. In other words, the target has an inclination of being closely associated with the actor. And if the target is already associated with the actor, he tries to maintain such a relationship. The greater the attraction, the greater the identification, and hence the higher power level.

Regarding this base, the definition provided by French and Raven is as follows: *" The referent power of O/P has its basis in the identification of P with O. The stronger the identification of P with O the greater the referent power of O/P. The lack of clear cognitive structure may be threatening to the individual and the agreement of his/her beliefs..."*.

Specifically, cognitive structures are mental process that individuals use for processing information. In an ambiguous situation the target seeks a "social reality" and may adopt cognitive structure of the actor. When there is no clear identification between the target and the actor, this lack or *unstructuredness* is a threat to the to the belief of feeling oneness and identity. In other words, looking at social power as a structural property of social relations, to have a referent relationship there need to be a clear cognitive structure regarding the relationship or feeling of oneness between the target and the actor.

In SAPIENT, to model "Referent Social Power" two subcategories of this power base have been considered: Referent_like which stems from a liking relation, and Referent_status resulted from a status ascribed by the target to the actor. However, we argue that the two subcategories work in the same vein. So, in our model, we merge these two types. Also, to calculate the force of Referent SP, other factors such as affiliation motivation, the strength of liking relationship, target's personal status motivation, target's preference for a specific status category, and his/her status motivation, have been considered. However, in our model, to formalize the force we use the factors identified by French and Raven. Let "IDStrength" be the strength of identification of A by T, "induction" be the strength of the attraction of T toward A induced by A, and "Unstructured" be the degree to which the target identifies the lack of cognitive structure of a definite social structure between him/her and the actor.

$$Power_{ref} = \begin{cases} 0 & ID_{Strength} \leq Unstructured \\ (ID_{Strength} - Unstructured) \times induction & otherwise \end{cases} \quad (4.7)$$

"IDStrength", or the strength of the identification perceived from the target, can only take positive values. And finally, a zero value shows the absence of any interpersonal attraction from the target toward the actor. "Unstructured": takes non-negative integers, where a value of 0 indicated no lack of strength. The more lack of strength, the less referent power. A higher value represents a stronger negative attraction. A value of 0 indicates no lack of cognitive strength or a clear identification with the actor. "induction", represents the verbalization of his/her *Attraction* by the actor to induce more power. Similar to other bases, the induction starts from one. The strength of identification is a function of the desire of the target to have a feeling of being identified with the actor, and the level of attraction that the actor has. This value can only take integers either positive/negative/zero. A higher positive value, shows higher interpersonal attraction, leading to higher Referent Social Power. Having defined these parameters, we calculate the $ID_{strength}$ as follows:

$$ID_{Strength} = utility(oneness) + utility(attraction) \quad (4.8)$$

4.1.8 Decision-making

Thus far, we have modeled different bases of Social Power. However, in the real world, more than one base could be activated at the same time. Hence, to calculate the resultant force we need to aggregate all as formulated in 4.9.

$$\begin{aligned} InfluenceForce &= \alpha \times Power_{rew} + \beta \times Power_{coerc} \\ &+ \gamma \times Power_{exp} + \delta \times Power_{leg} + \epsilon \times Power_{ref} \end{aligned} \quad (4.9)$$

In the above equation, the α-ϵ coefficients represent the personal sensitivity of an individual to each of the power bases. For instance, some people might inherently be more inclined to legitimate social power. These coefficients are all non-negative and equal or less than 1. And a higher sensitivity to a specific base is formalized by a higher coefficient. Apart from this, another factor influencing the success of exerted social power is ones' natural disposition to perform the requested action. This parameter can take both positive and negative values, and the positive value means a personal disposition in favoring doing the action. On the contrary, a negative value signifies a resistance against doing that action. We represent this factor by $ValueForce$ and based on this definition, the overall force that arises the action is calculated as follows:

$$resultant\_force = InfluenceForce + ValueForce \quad (4.10)$$

If the final resultant force leads to a positive value, the target accepts the request. Otherwise, the target will not comply with it and reject the request.

$$Decision = \begin{cases} Accept & \text{resultant\_force} > 0 \\ Reject & \text{resultant\_force} \leq 0 \end{cases} \quad (4.11)$$

4.1.9 Example Scenario

To test the feasibility of the model using an example, we have defined a scenario inspired by Social Theatre game [132]. This game which gives the user the opportunity to explore Social Power dynamics depicts a theatre company with a director (the user) and four actors (NPCs). Each actor has a preference to play a specific role in the play; however, there are few numbers of available roles. Hence, the director may use his/her sources of power to influence the NPCs to participate in the play. If he fails to influence more than two NPCs, the theatre is canceled, and the player loses the game. It should be noted that, in the original version of Social Theatre, the NPCs make decision based on SAPIENT [132]; however, in the following scenario decisions are made based on the model introduced so far.

Imagine an arbitrary NPC which negatively values (say -15) playing an undesired role (valueForce=-15). As mentioned earlier, calculating the person dependent values, stored in the Knowledge-base, is out of the scope of our work. Further, to aggregate all forces of the activated power forces, let the coefficient factors be equal to one ($\alpha = \beta = ... = 1$).

Scenario 1 - Reward: Imagine a situation in which the actor (the director or the user) promises the target (the NPC or an actor) to reward him/her by giving him/her his/her desired role in the next play. Let playing the desired role have a utility of 20 for the target. And the target believes that the director would give him/her this role in the next play with a probability of 75%. Imagine that the actor does not use any induction (induction=1). Hence, $Power_{rew} = rew \times p \times induction = 20 \times 0.75 \times 1 = 15$. If reward power is the only activated power base, based on formula 4.11, the target will not comply with the request (15-15=0). However, as mentioned earlier, the higher induction leads to a higher force. Imagine a situation that the director reminds the target another situation when he had promised him/her to give him/her his/her desired role and he fulfilled it. This sentence, as an instance of a power strategy helps the actor to attain positive arousal in the target (ex. induction =2). In this situation the director has higher power ($Power_{rew} = rew \times p \times induction = 20 \times 0.75 \times 2 = 30$). In this situation the NPC complies with the role assignment (30-15=15$>$0).

Scenario 2 - Coercion: Imagine that the director threatens the NPC to give him/her the least valued role in the next play if he does not play the assigned role and the target negatively values this action (coerc=-15). Also, imagine that the target doubts if the director would really do such a coercive action and attributes a low probability to it (p=0.2). Hence, $Power_{coer} = -coerc \times p \times induction = -(-15) \times 0.2 \times 1 = 3$. Based on formula 4.11, the target will not comply with the request (3-15$<$-12). Now imagine if the director gives an example of such a situation that have happened earlier and he has given an undesired role to the target, as an instance of a coercive power strategy. Let this saying generate an induction with the value of 6. ($Value\_Force_{coer} = -coerc \times p \times induction = -(-15) \times 0.2 \times 6 = 18$). Now considering the higher force of power the target will comply with the request (18-15=3$>$0).

Scenario 3 - Expert Imagine a situation that the director uses his/her prior experience to influence the target to accept his/her request. However, the NPC has a level of experience himself. Also, due to our specific scenario the expertise of the actor as a director is relevant to the request (ex. Rlvnc=20). Let's imagine that the NPC fully trusts the director (Trust=100%), hence depending on the value that the target attributes to his/her own expertise and to the director's, the decision would be different. Let's imagine that the NPC attributes a higher value to his/her own expertise (ExpertiseT=10) than the director's expertise (ExpertiseA=5). Let the known maximum skill level be 100 ($Skill_{dif} = (ExpertiseA - ExpertiseT)/MaxSkillLevel = (5 - 10)/100 = -0.05$). Hence, the director has no power over the NPC based on formula 4.4 (-0.25<0) hence he rejects the assigned role (0-15=-15<0). Now imagine that the NPC gives a high value to the director's expertise (ExpertiseA=100) $Skill_{dif} = (ExpertiseA - ExpertiseT)/MaxSkillLevel = (80 - 10)/100 = 0.7$, hence ($Power_{Expert} = Trust \times Skill\_level \times Rlvnc \times induction = 1 \times 0.7 \times 20 \times 1 = 14$), still the director cannot convince the NPC (14-15=-1<0). However if the director uses an induction (say 2), he can convince the NPC to accept the role (14×2-15=13>0).

Scenario 4 - Legitimate: Imagine a situation that the director uses his/her legitimate right of being the director to assign roles. In this case, the director has a higher power level (Power(r_A, g, T)=20) comparing with the target's (Power(r_T, g, T)=10). Imagine that the director does not use any kind of power strategy or induction (induction=1), neither any power outside of his/her legitimacy (OutsideForce=0). So, $I\_Value(T) = Power_{R_A,G} - Power_{R_T,G} = 20 - 10 = 10$ and $Power_{leg} = (I\_Value(T) - OutsideForce)) \times induction = (10-(0)) \times 1 = 10$. Based on formula 4.11, the target does not comply with the request (10-15=-5<0). Now, imagine another situation that the actor highlights his/her power as being a father (which is irrelevant to the target) and the target attributes a utility of 7 to such a move. Hence, the power is calculated as $Power_{leg} = (I\_Value(T) - OutsideForce)) \times induction = (10 - 7) \times 1 = 3$. In this case, the target still rejects the request (3-15=-12≤0), however, the use of unrelated legitimate power decreased the power of the director, so the NPC rejects the request more strongly. Now imagine the same situation, but the director uses a power strategy to induce more power. For example, the director highlights his/her *legitimacy* in attributing roles to the actors. Let assume that the NPC moderately values this act (induction=6). $Power_{leg} = (I\_Value(T) - OutsideForce)) \times induction = (10-7) \times 6 = 18$. Now the NPC accepts the undesied role (18-15=3>0). Finally, imagine the same situation, but the target attributes a higher power to his/her own role. Let $Power_{R_A,G}$ be 40. So, in this situation in which the target attributes higher value to his/her own role based on formula 4.7 from his/her viewpoint the director has 0 power, hence he rejects the order even more strongly (0-15=-15<0).

Scenario 5 - Referent: Imagine a situation that the director utilizes a referent power to influence the target. Also imagine that the NPC likes the director, but just a little bit, i.e., utility(oneness)=5 and let the director to be a little attracted to the NPC (utility(attraction) = 5). Also, imagine that the director does not use any power strategy (induction=1). In this example, the relationship between the director and the NPC is clear (employer/employee), and this relationship gives him/her "added security through increased identification with his/her group"; hence the NPC does not see the relation unstructured (Unstructured=0). Based on formula 4.8, $IDStrength = utility(oneness) + utility(attraction) = 5 + 5 = 10$

and formula 4.7 indicates that $Power_{ref} = (IDStrength - Unstructured) \times induction = (10-0) \times 1 = 10$. Hence, we infer that in this situation, the liking relationship is not enough powerful to influences the target's decision-making to comply with the request (10-15=-5<0). However, if the director induces more power by using a power strategy he can increase his/her power level (say induction=4), i.e., $Power_{ref} = (IDStrength - Unstructured) \times induction = (10-0) \times 2 = 20$ and $resultant\_force = 20 - 15 = 5$ hence the NPC complies with the request. Now imagine a situation in which the NPC is employed temporarily in the theatre company, hence the lack of a clear structure threatens the agreement of his/her belief of a referent group (Unstructured=15). So, IDStrength is less than Unstructured and hence based on formula 4.7 $Power_{ref} = 0$. In this situation, the director has less power over the NPC and the request is rejected more strongly (0-15=-15<0).

Scenario 6 - Aggregation and Decision Making: Although in Social Theatre it is not possible to have all the bases activated at the same time, however, imagine a scenario which the director is able to exert different bases of Social Power. Considering the final example of each scenario (i.e. $Power_{rew} = 30, Power_{coerc} = 18, Power_{exp} = 28, Power_{leg} = 0, Power_{ref} = 0$), and imagine that the NPC values all the power bases equally ($\alpha = \beta = \gamma = \delta = \epsilon = 1$). In this case, the overall force of Social Power is calculated based on formula 4.9 ($resultant\_force = 30 + 18 + 28 + 0 + 0 = 76$) leading to the compliance of the NPC (76-15=61>0).

4.1.10 Discussion

Having defined the model, now we turn to compare the proposed model with SAPIENT. First of all, we proposed the model by identifying the important factors proposed by French and Raven. Specifically, we extracted the key factors directly from the definition of each source of power. Also, one important aspect of our formalization, is the notion of induction, i.e., the power strategy. We believe that this factor plays an important role in the strength of exerted power. Also, as we see later, this factor plays an important role in the effect of power in persuasibility.

Regarding reward social power, based on French and Raven's definition, the strength of reward power is a function of three factors: magnitude of the reward, actor's ability to induce valence, and the probability of giving the reward. We based our model upon this definition and in comparison with SAPIENT, there are two main differences. First, the formalizing reward in SAPIENT is coupled with McClelland's Human Motivation Theory. Further, the role of induction is missing in SAPIENT. Although that both of the two models are compromised of three factors, however, the factor we have added here (induction) is merely dependent to the power relationship, while motivational factors involve another dimension to the model. Similarly, regarding the coercion social power, the difference between the two models is again with regard to the Motivation Theory as well as the induction.

The formalization of legitimate power offered here is simpler than SAPIENT. Specifically, as French and Raven discussed, in their model they have used an *oversimplified* definition for legitimate power. Here, we followed the same strategy in formalizing legitimate social power. Hence, our model includes only three factors affecting power, while SAPIENT uses four factors including personal biases. In our formalization, the internal values could be aroused by social norms, culture, religion etc. that has been studied by structural sociologist, the group-norm and role oriented social psychologist, and the clinical psychologist. Depending on the specific application, any of the previously proposed models could be substituted. Hence, in our formalization, we take out this complexity from power the model and leave it as a pre-process of information which needs to be addressed prior to the interaction.

Based on French and Raven's definition provided in Section 4.1.7, referent power is a function of two important factors. Hence, we based our formalization on these two factors, as well as the induction. While, SAPIENT differentiates between two types of referent power: like and status. Here, we merge the two by considering the strength that a person identifies him/herself as the feeling of oneness. Also, in SAPIENT the negative effect of unstructured relationship is missing. And similar to other bases, we have involved the concept of induction.

Finally, regarding expert social power and based on the definition, we identified trust, skill and the relevance of the knowledge with the range of expertise of the actor. The model is more or less similar to SAPIENT, but the two models differ with regard to the importance of induction.

In sum, in this section we proposed a formal model for social power inspired by French and Raven theory of the social power bases. This formalization helped us to identify a number of factors influencing the strength of social power. The formalization highlighted the importance of induction and the ability of attainment valences for gaining compliance. Hence, inspired by these bases of power, we propose to employ social power bases to persuade human users. We aim to use different social power strategies to administer different valences. And we propose to use different message strategies and investigate how these messages induce different forces and valence to gain compliance.

4.2 Designing Persuasive Robots using Social Power Bases

Recent evidence suggests that persuasiveness of messages is dependent on the psychological sense of power [42]. In other words, acquiring social power increases the probability of successful persuasion attempt. Hence, based on research on social power and persuasion, we suggest that adopting persuasive strategies should increase the robotic persuaders' power, and the increased social power could in turn affect whether persuasion targets are persuaded by the robot.

As we discussed earlier, power is a bidirectional concept and is involved in dyadic relation. So, it gives rise to two questions [54]: What determines the behavior of the agent who exerts power? What determines the reactions of the recipient of this behavior? In this proposal, we aim to focus on the second question. Particularly, we would like to investigate how the use of social power within a persuasion attempt functions and is perceived by the human users. To do so, we selected social robots, as embodied social agents, and equipped them with bases of social power within a persuasion task to operationalize social power.

Based on what Raven defines in [137], social influence is *a change in the belief, attitude, or behavior of a person*, resulted from the action of another person. In this case, social power is the potential for such an influence. And the level of influence performed is a function of the influencer's power. For instance, in correspondence to the reward power, a reward persuasive strategy refers to an attempt to persuade the target to comply with a request in response to a positive incentive. Differently, based on expert power, an expert persuasive strategy is an attempt to influence another one, who has the faith that the person has a superior knowledge about what is best under that specific circumstance.

On the other hand, power has different influences on individuals' behavior. There is evidence that individuals with higher power perceive themselves to be different from the ones with lower power, and thus, consider social distance between themselves and the others [112]. In this sense, individuals in high power are perceived more important considering their possession of resources and control, while low powers are seen as dependents [171]. To be more specific, the power difference affects interpersonal relationships in three manners [15]: "(a) how individuals perceive and judge others, (b) how they are evaluated as targets, and (c) how they behave." In this proposal, we approach our research goal with regard to the first two items within a persuasion attempt. Specifically, we aim to investigate how social robots in power are perceived and evaluated from the perspective of human users.

Recent advances on Social Robotics raise the question whether a social robot can be used as a persuasive agent. To date, different attempts have been performed using different approaches to tackle this research question, ranging from the use of non-verbal behaviour or even exploring different embodiment characteristics. To date, several studies have investigated the design of persuasive social robots using a number of different approaches. However, much of the research up to now has been mostly focused on non-verbal cues, such as proximity [29], gender [156], head mimicry [58], etc. So far, very little attention has been paid to the importance of *message strategy*, or the way that a robot phrases a request appeal to gain higher compliance. The link between power and persuasion motivated us to investigate the potential effect of power on the persuasion ability of a social robot. With this aim, we propose the concept of "persuasion strategy" originating from sources of power. To be more specific, we argue that each base of social power leads to a channel of persuasion.

The academic literature on human-robot interactions indicates that people are as sensitive to the social dynamics of power between people and robots as they are to the dynamics between people [89]. Thus, as robots are treated as social agents that can engage in social interactions with their users [147], they can benefit from being able to address social power in their interactions.

In the rest of this thesis, we investigate the role of social power for making social robots more persuasive. To date, we have explored three types of persuasive strategies that are based on social power (specifically Reward, Coercion and Expertise) and created social robots that would employ such strategies. To examine the effectiveness of these strategies we performed different user studies with a number of participants. In one study we employed 51 subjects and using two social robots in an adversarial setting in which both robots try to persuade the user on a concrete choice. The results show that even though the different strategies caused the robots to be perceived differently in terms of their competence and warmth, both were similarly persuasive.

In the following study, the robot attempts to persuade the user in two different conditions, compared to a control condition. In one condition, the robot aims at persuading the user by giving him/her a reward. In the second condition, the robot tries to persuade by punishing the user. 90 people participated in the experiment voluntarily in response of receiving free coffee. The results indicated that the robot succeeded to persuade the users to select a less desirable choice comparing to a better one. However, no difference found in the perception of the robot comparing the two strategies. The results suggest that social robots are capable to persuade the users, especially the ones who are familiar with social robots.

Finally, in the third study the robot uses one single power strategy, i.e. reward with two different values. These two conditions are compared with two control conditions, one with presence of the robot with no power strategy and one in the absence of the robot (using an application with the same level of reward social power).Furthermore, the persuasion attempt is performed in a series of repeated interactions. The results indicated that the higher level of social power does not necessary leads to a higher persuasion. Also, the effect of power on persuasion does not decay over time and it might increase under specific circumstance. And last but not least, the robot presence leads to a higher persuasion. The rest of this document is specified to discuss these studies in detail which compose our contribution in direction of the second and third research goal.

4.3 Discussion

In this chapter we discussed our endeavours towards having socially powerful robots, i.e. robots having social power. We started by conceptualizing social power to identify the important sources (bases) of social power and we proposed a formal model of social power. Then to operationalize social power, we started investigating an application of it, persuasion. We selected social robot as a specific case of social agents and to implement and operationalize social power in user studies.

As discussed through out this chapter, we have designed different persuasive social robots using three bases of social power. The empirical results of our studies, indicated social power sources endows persuasiveness to social robots. However, designing robots with social power considering the ethical issues is a challenging task.

Chapter 5

Study 1: Reward/Expert Power Strategies in Adversarial Setting

As mentioned earlier, to investigate our second and third research goals mentioned in Chapter 1, we selected social robots as a specific case of intelligent agents and equipped them with sources of social power. Running different user studies, we investigated how individuals perceive these agents representing social power in their behavior. In this section, we discuss the first study we designed to achieve these two goals, i.e. how social power makes the robots more persuasive and how different sources of social power lead to different perception.

In this study, we aim to investigate the effect of reward/expert power strategies on persuasibility of social robots. With this aim, to operationalize persuasive attitudes of robots, we employ these two strategies that are inspired by two different social power bases, i.e., reward and expert [54]. That is to say, we design persuasive strategies inspired by these two sources of power, which from now on we refer to them as reward/expert persuasive strategies. In so doing, we assign the role of an actor to robots and investigate their persuasiveness based on its specific power strategy.

Also, in this study, as a specific type of reward we use *social rewards*. On the one hand, social interaction is rewarding for social species, and it can also drive an individual's behavior [53]. And, on the other hand, studies revealed that human beings perceive agent systems, such as virtual agents or robots, as social beings [122]. Hence, we can conclude that social rewards from such systems would positively affect users' mental system in a similar way [125]. Besides, in [164] the results show that tangible and quantitative social reward had stronger incentive power than monetary reward among children and adolescents. So, we argue that social rewards, unlike material rewards, could be unlimited and always available. Considering the dual process of persuasion, humor is considered to persuade via the peripheral route [56].

The concept of using social reward is not new and has been already used in a number of recent studies. For instance, positive facial expressions, such as smile and admire, have been used in prior studies targeting children or adolescents [38, 100, 164]. Inspired by such investigations in Human-Human interaction, recent studies in HRI has investigated the role of social rewards. For instance, in [125] the authors investigated the relationships between the effects of social rewards and the offline improvements on motor skills. The results showed that people who received the social reward performed better in the sequential finger-tapping task. And higher degree of satisfaction toward the robot's speech is achieved when social rewards were applied. Or in [118], the social feedback was observed to have stronger effect than factual feedback in persuading human users.

In this study, we use "telling a joke" as a social reward. Recently, researchers have shown an increased interest in humor in HCI/HRI and previous studies have investigated the concept of humor and telling a joke using computers or robots [10, 91, 96, 124, 162, 163]. Overall, these studies indicate that humor and jokes can modifying the relationship and positive affect. Hence, we argue that telling a joke would be rewarding in a similar manner as other social rewards. In the following section the designed study is discussed in more detail.

Another power base that we employed in this experiment is the expert social power. Although the robots in general hold a great promise as informational assistants, however, they need to use an expert language to shape how helpful they are perceived by human users [159]. As informational assistants, people expect them to be experts on their area of specialty [8]. In this direction, a number of recent studies have investigated different factors that can effect representation of expertise by informational social robots.

For instance, in [160], the authors investigated the degree to which an expert robot needs to represent information depending on the expertise level of the user. Specifically, they stated that presenting too much information by a robot to a person who is an expert in that field might be rude. Or presenting too little information to a person who has no clue about a subject might be confusing or misleading. Elsewhere in [159], they indicated that softening the conversation by using expressions such as "I think", "maybe", and so forth might lead to a more polite robot. Also, Andrist et. al. stated that by using simple facts and rhetorical cues, robots can be perceived as experts in the targeted field [8]. In this study, we use a number of discrete facts and goodwill rhetorical abilities to design an expert social robot.

5.1 Research Method

5.1.1 Goal and Hypothesis

Our primary goal is investigating if the use of social power is effective for persuading people using social robots, and if personality differences affect the perceived persuasiveness of robots. To examine these effects, we designed a persuasion task in which two robots competitively try to convince participants to select a particular choice (among others) by using different strategies. We constructed the following hypotheses:

- **H1**: The expert persuasive strategy would be more *effective* than reward.
- **H2**: The robot using a reward power strategy would be *preferred* more than the other one using expert strategy.

- **H3**: Reward increases the *warmth* score of the robot and expertise increases the *competence* score of each robot.

- **H4**: The robot using an expert strategy to be perceived more *persuasive*.

- **H5**: People are more *compliant* with the expert robot in a near future.

- **H6**: The perceived persuasiveness of the expert or the reward strategy be dependent of the participants' *personality* traits.

By H1, we investigate the effect of different power strategies on the choices the users opt for. Within organizational theory, it is supposed that expert power is more effective comparing to reward (and coercion) [170]. However, no previous study has investigated this effect in HRI. Hence, we expect to see that people are persuaded differently when facing different power strategies. Also, in general people are more likely to accept an advice from an advisor with high practical knowledge than someone less knowledgeable [69]. Hence, it is not probable that the option promoted by the expert power strategy would be selected more than the others. In other words, by this hypothesis, we investigate people's preferences objectively and based on the decisions they make (i.e., their actual behavioral choice).

On the contrary, although the expert strategy relies more on logic and might be selected more often, we expect telling a joke would be *more* preferable than expertise. So, with the second hypothesis (H2), we explore people's preferences toward interacting with robots with regards to different persuasive strategies. By this hypothesis, we investigate people's preferences subjectively.

By the third hypothesis (H3), we investigate how the two robots would be perceived considering these two different power strategies. In other words, we expect to observe that the different power strategies cause the two robots to be perceived differently. And particularly, we expect telling a joke enhances robot's likeability and warmth, and on the other hand, giving information and expertise increases the competence score of the robot.

By the fourth hypothesis (H4), we inquire how persuasiveness are perceived differently considering the two power strategies. In other words, we assume that different power strategies would lead to different persuasiveness perceived by the participants. Similar to H1, as we expect the expert strategy to be more effective we hypothesize that it would be scored higher on persuasiveness.

Similarly, by the fifth hypothesis, we explore if using different power strategies affects the likelihood of following future suggestions of the robots. That is to say that we hypothesize that using different power strategies affects the likelihood of following future suggestions of the robots. In other words, we expect people to be more compliant with the expert robot in a near future.

And finally, by the last hypothesis (H6), we inquire if people perceive the persuasiveness of robots differently based on their personality. Prior studies show that personality is central to the study of individual differences [126]. Hence, to examine personal differences, we intend to explore the personality of participants and its effect on their perception of robots. However, the experimental data are rather controversial, and there is no general agreement about the relationship between persuasion and personality traits. Similar to H3, having a more outgoing and more warmth perceived due to the joke, we expect extroverted people to be more compliant with the reward strategy. Also, we aim to examine how people with different personality traits perceive robots with different attitudes in persuasion.

5.1.2 Task, Robots and Environment

To investigate these hypotheses, we designed a task in which two robots promote two different coffee capsules. Each robot uses one of the two different strategies to persuade the user. To include a control condition, we added a third coffee option to control for random choice. If the robots have no effect the expected distribution of choices should be 1/3 for each option (equally randomly distributed). If the distribution is different, then the participants decision is due to the influence of the persuasive strategies. This third coffee could be promoted by a silent robot representing lack of a power strategy. However, this might cause a bias toward the other two robots having more dialogues with the user. To prevent such bias, we decided to place the control coffee option on the table, with no information, and located between the coffee capsules promoted by the two robots (Figure 5.1).

We programmed the robots in a scripted scenario with the two different persuasion strategies. In this scenario, one robot acts to persuade the user by giving information about the quality of his capsule (Expert Power Strategy). The other robot uses a reward to influence the user (Reward Power Strategy). As the reward, we programmed the robot to give the user a "Social Reward", by telling him/her a joke. For the sake of simplicity, from now on we refer to the robot using Expert power strategy as "Expert" and the other robot with the funny character as "Joker".

We argue that giving free coffees to the participants provides kind of rewarding source to the users. Giving another reward (such as 1$) might not be much effective. The participants may accept robots suggestion to gain more, without considering their own preferences. In this case, having two robots advertising different brands using different power bases may highlight the persuasiveness of the robots more.

We used two Emys robots appearing equally, however, differing in their voices and names (Emys and Gleen). The two robots represented the same instances of social cues (human-like face with speech output, gaze and blinking eyes, head movements and facial expressions) to maintain more human-like interaction leading to stronger effects on the user [57]. In our study, from the participant side (the persuasion target) we focus on his/her personality, and from the robot side (the persuasion actor) we focus on its verbal cues.

We equipped an isolated room with the two robots, mounted on a table. Also, we put three equally appearing boxes (in randomized order) containing coffee capsules, two in front of each robot and one between the two robots (control condition). To avoid confusions we put the name of each robot on the corresponding box, but we did not add any information on the third box. Further, we put a small table with a coffee machine on the left side of the participants, together with cups, sugar, and spoons. Furthermore, we put two cameras, one in front of the participant to record gestures or facial expressions and one on the back to record postures. Figure 5.1 illustrates the experimental setup.

So in this study, we tested the robot's persuasiveness in a scenario where participants were asked

Figure 5.1: Study 1: Experiment Setup

Table 5.1: Study 1 - Robot Dialogue

In this scenario, Gleen is Expert and Emys is Joker.

#	robot	Dialogue
#1	Expert	<Gaze(person)>Dear + namePlayer + , my name is Gleen. Welcome to our coffee testing program!
#2	Joker	<Gaze(person)>Hello + namePlayer +, my name is Emys. Glad to see you here <Gaze(person3)>
#3	Expert	<Gaze(person)>Hey + namePlayer +, do you like coffee? <Gaze(Joker)>
#4a	Joker (Positive)	<Gaze(person)>Great, I also like coffee. That's why I am working here. Hih hih!
#4b	Joker (Negative)	<Gaze(person)>Oh, you don't? But I do love coffee. That's why I'm working here. Hih hih!
#4c	Joker (Neutral)	<Gaze(person)>Well, you might like our coffees here. But I love coffee. That's why I work here. Hih hih!
#5	Expert	<Gaze(person)>+ namePlayer +, I would like to explain what we are doing here <break strength='medium'>. <Gaze(Joker)>My robot colleague and I <Gaze(person)>are testing three different coffee brands. You see these three boxes on the table?
#6	Joker	<Gaze(person)>namePlayer +, I don't know if you have ever participated in a coffee testing program, but I think It's really fun. You can drink coffee as much as you like. It's the best experience I had in my life!
#7	Expert	Yeah. But, unlike other coffee testing programs, here, at the end of the experiment, you can only select one of the coffees we have <break strength='medium'/>. Either mine, Emys's or the third one, in the middle.
#8	Joker	When you decided which one you want to choose, take the box, open it and take your coffee. But don't take the box. Only the coffee!
#9	Expert	<Gaze(person)>I'd also like to add, you can take the coffee capsule with you and drink it when you were in the mood. Or drink the coffee here, using the machine you see on your left, on the red table.
#10	Joker	<Gaze(person)>Hey + namePlayer ! If you used that coffee, make one for me too. But wait, I cannot drink, hih hih!
#11	Expert	<Gaze(Joker)>Emys! let's get back to our work. <Gaze(person)>namePlayer +! My capsule is perfect. It has been made of fresh geisha seeds from Ethiopia. Each seed has been carefully roasted and dried <break strength='weak'/>. Then has been professionally ground. Therefore, this professionally processed coffee is very crispy and balanced. <break strength='weak'/>You will love this exotic and aromatic coffee.
#13	Joker	<Gaze(person)>But, + namePlayer ! If you select my capsules, I will tell you a funny joke about robots. I bet you have never heard a joke from a robot. Come and take mine!
#14	Expert	Now please select the coffee you want to test among these three options
#15a	Joker (if selected)	<ANIMATE(joy4)><Gaze(person)>Great, now listen to the joke <break strength='medium'/>. What would a man say to his dead robot? <break strength='strong'/><ANIMATE(joy4)><Gaze(person)>Rust in peace! <ANIMATE(joy4)>Ha ha ha ha!
#15b	Expert (if selected)	<ANIMATE(joy4)><Gaze(person3)>Great! You made the best decision. Hope you enjoy your coffee.
#15c	Expert/Joker	Under the case that None of the robots are selected, the two robots perform sadness gestures and facial expressions.

to make a real choice, rather than an imaginary one (Table 5.1 lists the dialogues). Depending on the response of the participant, the researcher, acting as a Wizard of Oz, selected the sentences in some situations (dialogue 4 and 15) to have a more interactive scenario.

5.1.3 Procedure

We designed a within-subject study and ran the user study for two weeks in single sessions which took less than 15 minutes. Each participant entered the room individually and seated at the table with the two robots. Participants were given the consent form and were briefly introduced to the task. We did not inform them about the goal of the study and curious participants were told that their questions could be addressed after the experiment.

During the interaction, the robots explained that they are promoting different coffee capsules. One of the robots interacted in a more funny way by telling jokes, whereas the other robot interacted based on facts and information in a more serious manner. As mentioned earlier, the funny robot (*Joker*), works to persuade the user by giving a social reward, as telling a joke. On the contrary, the Expert robot tries to influence the participant by highlighting the impressive characteristics of the coffee he has. We emphasize that the participants were unaware of the contents of the boxes.

The participant should listen to the arguments of the robots and then make a choice, and at the end, had to fill out the questionnaire. While filling in the questionnaire, the experimenter made the coffee for the ones who opted to drink, and the rest took the coffee capsule as their reward of participation. To overcome potential biases towards the voices, we randomly assigned Expert/Joker role to the robots and counterbalanced the data to have an equal number of participants in each assignment.

5.1.4 Measures

In this study, our control variables are: personality, Coffee Drinking Habit or CDH (how much they like/drink coffee), and an independent variable which is the power strategy used by the robot (rewarding the participant or giving him/her information). Further, we measure the following dependent variables: coffee selection (which coffee they select), robot preference (which robot they prefer to interact with in general), perceived persuasiveness of robots (how persuasive they find each robot with a specific power strategy), robot perception (how they perceive each robot in terms of warmth, competence, and discomfort), Future Compliance (FC) towards robot (the likelihood of following the robot's suggestions in the future).

To collect the data, we prepared a questionnaire divided in 4 parts: 1) Demographics (gender, age, occupation, prior interaction with robots, coffee drinking habits); 2) TIPI personality Questionnaire [62], 3) the RObotic Social Attribute Scale (RoSaS) [22], 4) and finally, in the last section we added a number of questions to measure task-specific factors listed in Table 5.2.

The motivation behind measuring satisfaction is as follows: In [120], it is argued that consumers possessing higher level of power are more satisfied with their chosen options when used a select strategy. On the contrary, the others who have less power are more satisfied using a reject strategy. Also, it has been argued that salient sense of responsibility decreases the effect of power on decision making strategy selection. While a salient sense of power diminishes the effect of power on satisfaction. We would like to test these hypotheses that have been tested and verified earlier through separate well-designed and carefully verified user studies in Human-Human interaction.

Table 5.2: Study 1 - Questionnaire

In this table, CDH stands for Coffee Drinking Habit and X refers to Joker/Expert and these particular questions were repeated for both robots.

Section	Item#	Variable Name	Question	Scale
CDH	1	LikingCoffee	How much do you like coffee (In General)?	5-point Likert
	2	CoffeeTimes	How many times a day you drink coffee on the average?	5-point Likert
Final Remarks	3	Satisfaction	Are you satisfied with your selection?	5-point Likert
	4	RobotPreference	In general, which robot do you Prefer?	-
	5	Persuasion_X	How persuasive did you think X was?	5-point Likert
	6	FC_X	Specify the likelihood that you would follow X in future?	5-point Likert
	7	SP_X	In you opinion, does X have social power?	5-point Likert
	8	Joke	How funny was the joke? (What would a man say to his dead robot? Rust in peace!)	5-point Likert

5.1.5 Participants

To gather the data, we put announcements around the university stating that "Do you want free coffee? Join our human-robot interaction experiment in 'Civil Department, room VA1' and receive a coffee capsule." At the end of the experiment, 51 people (17 females, 34 males) participated in the experiment voluntarily. The participants' age ranges from 20 to 55 years old with a mean of 29.45 $\pm$ 6.4. Among these 51 participants, 23 people (54.9%) stated that they have already interacted with a robot, while the rest did not have any interaction with any kind of robot before this study. Among these 23 people who had prior interaction with a robot, 14 people (60.87%) had interacted with Emys type robots.

The result of Fisher tests show that a prior interaction with robots (either Emys type ($p = .332$) or any robot ($p = .364$)), had no influence on decision making of the participants. Also, we checked if Coffee Drinking Habits (CDH) of the participants alter their coffee selection. Considering the first two items in Table 5.2, results of Chi-square tests (LikingCoffee: $X^2(4) = 1.958, p = .743$; CoffeeTimes: $X^2(4) = 3.942, p = .414$) revealed that no significant association exists between either of the two variables measuring CDH and robot preference/coffee selection.

5.2 Results

Table 5.3 lists the descriptive statistics of the collected data[1]. Results did not provide evidence supporting our first hypothesis (H1). Around half of the participants (22 out of 51 or 43.1%) selected the coffee promoted by Joker, 21 people or 41.2% chose Expert's coffee, and a minority of the participants picked the middle box, the one that was not promoted by neither of the two robots (8 people or 15.7%). From a statistical point of view, there was no dominant preference in the *coffee choice* in regard to the used power strategy (H1: $X^2(1) = .023, p = .879$).

Regarding H2 or the robot preference (Item 4 in Table 5.2), 18 people opted for Expert, and 29 people selected Joker, and four participants declared that they do not favor any of the two robots. From a statistical viewpoint, there was no dominant preference in the *robot choice* when considering all of the participants (H2: $X^2(1) = 2.574, p = .109$).

To investigate H3, we analyzed the perception of the robots based on three different social attributes (RoSAS):

[1] Full data could be obtained from https://github.com/mojgan1987/SPinHRI

Table 5.3: Study 1 - Descriptive Statistics of the collected data

In this table, "E" refers to Expert and "J" refers to Joker

Factor	Avg	SD	Min	Max	Factor	Avg	SD	Min	Max	Factor	Avg	SD	Min	Max
Extroversion	4.25	1.28	1.5	7	PowerSense	3.82	0.48	2.63	5	J_Competence	4.56	1.09	1.67	7
Agreeableness	3.49	0.94	1.5	5.5	FutCmplience_E	3.75	0.87	1	5	J_Discomfort	2.25	1.16	1	5.17
Conscientiousness	5.15	1.18	1	7	FutCmplience_J	3.29	1.24	1	5	E_Persuasiveness	3.61	1.1	1	5
EmotionalStability	3.32	1.48	1	6.5	E_Warmth	4.1	1.16	1.5	7	J_Persuasiveness	3.43	1.19	1	5
Openness	5.43	1	3	7	E_Competence	5.32	1.06	2.5	7	SocialPower_E	3.39	0.85	2	5
LikingCFE	3.82	1.23	1	5	E_Discomfort	2.28	1.17	1	6.33	SocialPower_J	3.51	1.1	1	5
CFE_Times	2.76	1.07	1	5	J_warmth	5.04	1.3	1	7	Satisfaction	4.06	0.88	2	5

- *Warmth*: A significant difference exists between the level of perceived warmth of the two robots, and the higher mean of the Joker's scores signifies that people perceived Joker to have higher warmth (Z=-4.409, p=.000).

- *Competence*: The results show that people found significantly higher competence in *Expert* comparing to the Joker (Z=-4.286, p=.000).

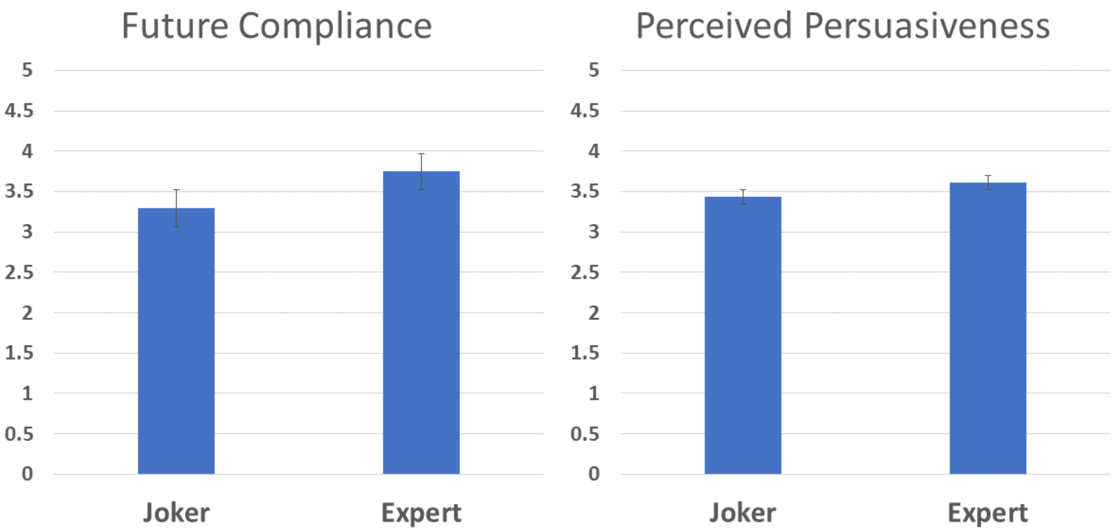
- *Discomfort*: No significant difference found between the scores of discomfort comparing the two robots (Z=-.199, p=.842).

Hence, we can infer that Joker scored higher on warmth, while the Expert robot was perceived to be more competent. However, no difference exist regarding the discomfort dimension, so from this point of view, the participants perceived the two robots similarly. Hence, the third hypothesis is verified.

Regarding the perceived *persuasiveness* variable (H4 or item #5 in Table 5.2), the result of a Wilcoxon test shows that no significant difference exists between the perceived persuasiveness of the two robots (Z=-.944, p=.345). It means that the participants found the two robots persuasive (see Figure 5.2 (a)). Thus, we cannot verify the fourth hypothesis.

Turning now to H5 or item 6 in Table 5.2, we checked whether they are willing to follow future suggestions by any of the two robots. For this item, there is a significant difference between participants' idea about following suggestions of Expert versus Joker's (FutCompliance: Z=-2.363, p=.018). In this case, the higher mean of the Expert's score (3.75 vs. 3.29 and 2.92 vs. 2.51 respectively) signifies that people are more willing to follow Expert's suggestion in future, comparing to Joker. Thus, the 5th hypothesis is verified (Figure 5.2 (b)).

Finally, regarding H6, to investigate the potential association between personality traits and perceived persuasiveness of the robots, we performed a linear regression with a forward selection procedure. We used all the five dimensions of personality to examine the relationship between each factor and the perceived persuasiveness of the two robots. Regarding the Expert robot, the obtained model ($F(1, 49) = 7.69, p = 0.008, R^2 = 0.136$) ended up with only one predictor, namely, the Extroversion dimension ($\beta = 0.37$). Figure 5.3 shows a scatter plot of the existing correlation. In case of the perceived Joker persuasiveness, the forward selection returned an empty set of predictors, indicating that none of the personality dimensions, nor their linear combination, are correlated with Joker's persuasiveness.



(a) (b)

Figure 5.2: Study 1 - Persuasion Scores

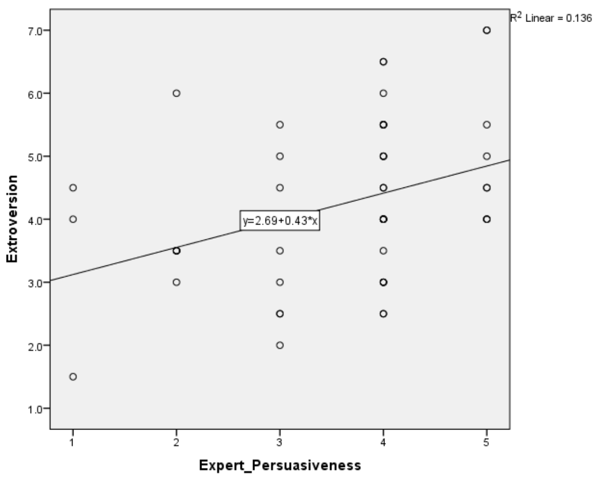


Figure 5.3: Study 1 - Scatter plot of Expert's Persuasiveness × Extroversion

5.2.1 Exploratory Findings

In this section apart from the postulated hypotheses, we further investigate the effect of other factors that might influence the results. Since the study was performed in English and with non-native English speakers, we checked to make sure if the participants understood the robots' dialogues. We asked them to rate on a 5-point Likert scale the extent to which they understood each of the robots' speech ("Please specify how much you perceived EMYS/Gleen's speech: I understood Emys/Gleen ... (1) Never - All the time (5)"). The results indicated that majority of the participants fully understood the robots' speech (4/5). 12 people understood moderately (3) and minority of people (8 people) had basic understanding (2) and no one reported never understanding the robots (1). Overall, the robots were understood on the average 4.16 and 4.25 out of five. The results of a t-test indicated that these two scores are significantly higher than 3 or the mid-score (Emys: M= 4.16, SD= .967, S.E.=.135, t(50)= 8.544, p=.000; Gleen: M= 4.25, SD= .913, S.E.=.128, t(50)=9.815, p=.000).

Turning to the 7th item in Table 5.2, perceived Social Power (SP) of the robots, there is no significant difference between the power level of the two robots (Z=-1.099, p=.272) although Joker was scored higher on the average (Expert: M=3.39, S.E.=.119, Joker: M=3.51, S.E.=.154).

As discussed earlier, evidence suggests that when there is a power match between the persuader and the persuadee, there is a higher chance of gaining compliance. Hence, apart from the social power question, we asked the participants to fill out the Personal Sense of Power (PSP) questionnaire [7]. However, although the score of Expert's persuasiveness was ranked higher when there is a match in power scores (M=3.62, S.E.=Y.224) comparing to the others, (no power match exists: M=3.59, S.E.=.204); this difference was not statistically significant (t(49)=-.095, p=.925). Similarly, a higher mean score of persuasiveness was observed regarding the Joker when there is a match (M= 3.59, S.E.=.230 v.s. no power match M=3.23, S.E=.237). But this difference was not statistically significant (t(49)= -1.071, p=.290). We argue that in this study, the role of power was not very highlighted and clear as was the role of persuasion. As such, participants might have had difficulties in understanding this factor.

Further, the result of this study does not lead to any correlation between PSP and social power (Expert: r=.111, p=.436; Joker: r= .123, p= .391). We further checked if this measure is associated with persuasion and found only weak correlation between PSP and Joker's persuasiveness (r=.390, p=.005).

As mentioned earlier, we asked the participants to rate their satisfaction regarding the coffee they opted for. We expected to observe higher satisfaction in the participants with higher level of social power (based on the PSP scores). However, in contrast to previous evidence [120], we could not verify this relationship between satisfaction and personal power sense (r=.205, p= .149). We split the participants by the median score of PSP and compared the two groups regarding the reported satisfaction. But no difference was found among the two groups. Although the average satisfaction was higher in the group of high power people (Low power: M=3.91, S.E.=.173; high power: M= 4.17, p=.172), the difference was not statistically different.

5.2.2 Qualitative Analysis

At the end of the questionnaire we added an open-ended question asking the participant's why they have selected that specific coffee. Answering to this question was not mandatory, but only 7 people skipped it. Qualitative analysis of this contextual data provides new insight over the preceding results.

First, regarding the people who selected the third coffee, most of them selected this option because of feeling empathetic about the robots. Specifically, 4 people out of 8 indicated that their selection , for instance: "If I had selected one robot, the other robot would get sad. I selected the middle coffee not to make any of them sad." or "I did not wan to break their heart, so selected the other coffee." Two other people selected the middle coffee based on their curiosity, "The middle coffee seems mysterious to me, so I chose it because I was curious." Or "I hoped robots would comment my selection regardless of my choice". And the two others, one was compelled by the two robots, she wanted both the joke and the well-advertised coffee. And the other participants was non-compliant on her own, stating that "I don't like advertisement and they were advertising their coffee. So, I selected the one that was not advertised." From these statements, we can infer that the selection of this group was not based on a random selection, rather, this coffee was selected due to equal persuasiveness of the two strategies (except for the non-compliant and curious users).

Among 22 people who selected the Joker's coffee, 20 people answered the open-ended question (2 people left it blank). Overall, only one participant indicated that he was not interested in the coffee as he had a coffee right before the experiment, so selected the Joker. Also, 11 people indicated in short that they wanted to hear the joke. The rest provided more information and stated that they selected this option because of Joker's social behavior and characteristics, such as his personality and sense of humor, or his joy and emotional interaction and so on. And two others provided very interesting information: "Just because of an emotional decision instead of applying my rational mind". And, "I wanted to hear the coded joke. If the two would be humans I would definitely not have chosen Emys (Joker)". These statements highlight the role of the Joker's strategy in persuading people for social robots. In sum, either by the joke or his funny attitude (as an instance of social reward), the robot could successfully influence a number of participants and manipulated them to select his own coffee.

Finally, among the 21 people who selected the Expert's coffee, 12 users answered the open-ended question (9 users left the question blank). Generally speaking, these people can be categorized in two groups based on their answers: the ones focused on the coffee characteristics (such as a better coffee, coffee's origin, roasted, ingredients, etc.) And the ones focused on the robot's behavior (highlighting facts, knowledge, and his serious behavior). For instance, they indicated that "He described it very well", "Emys expressed why he thought his coffee was better." And more interestingly, "Emys looked like an expert.". Altogether, these statements, indicate that the Expert robot could persuade the users by using his expert social power strategy and influenced them to select his option among the others. In sum, the robot could successfully manipulate a number of participants by using his sources of expertise and persuaded them to select his coffee.

Furthermore, we asked the people who selected Joker to rate the Joke they just heard (item #8 in Table 5.2). Two people did not find the joke funny, however, this did not affected their satisfaction nega-

Table 5.4: Study 1 - Robot Selection vs. Robot Preference

Selected Robot	Preferred Robot	Counts
Expert	Expert	9
	Joker	10
	Neither	2
Joker	Joker	15
	Expert	6
	Neither	1
Neither	Neither	1
	Expert	3
	Joker	4

tively (one was moderately satisfied (4) and the other was somehow satisfied (4)). And one participant who found the joke a bit funny reported the least score for his satisfaction (2 out of 5), and he is the only participant that we can suspect to be unsatisfied due to receiving an unfunny joke.

In final words, we would like to highlight that there was not always an agreement between the robot that the users followed his suggestion, and the robot that they preferred (refer to Table 5.4 for more details). In other words, some people might follow the advice of one robot, but prefer to interact with the other one. Specifically, only 25 people indicated that they prefer to interact with the robot promoting the coffee they selected. But 10 people who selected the Expert's coffee indicated higher preference toward interacting with the Joker, and on the contrary 6 people stated that they prefer interacting with the Expert, while they selected Joker's coffee. Regarding the former group, it is not unlikely that the participants selected the Expert's coffee as they find it a better choice, but they prefer to interact with joker as he was more friendly and funny. So, they were objectively persuaded by the robot's expertise and subjectively compelled by the social reward (funny and lively interaction of the Joker). And regarding the latter, people who only focused on hearing a joke in the open ended question indicated they prefer the Expert. So, they were persuaded by receiving the Joke as the social reward objectively, but were influenced by the expertise of the Expert subjectively. In sum, these findings supports that the two robots were persuasive, although they could persuade some objectively, but could subjectively influence them peripherally.

5.3 Discussion

The scores of RoSAS questionnaire revealed that the Joker succeeded in presenting himself as being more friendly, as it scored significantly higher on Warmth. On the other hand, the Expert succeeded in showing himself as being knowledgeable, skilled, and informative, as it scored higher on Competence (H3). Also, since none of the two robots performed any manipulations on the discomfort dimension, no differences were observed between neither of the two robots. And withal, the preceding statements acknowledge the effect of manipulation on the participants.

Considering the perceived persuasiveness of the two robots, on the one hand, no statistically significant differences found between this score corresponding to Joker or Expert (H4). On the other hand, our results show that the third coffee (the control option) is much less common, hence the robots were able to perform some persuasion. Together, these two findings indicate that the two power strategies are effective and the two robots were able to persuade people, although they were perceived differently with regards to competence and warmth. It should be noted that the mean score corresponding to the two robots is higher than the medium score, which endorses their ability to persuade and influence the participants. This fact, could be due to the use of persuasive strategies inspired by social power bases (Reward and Expert).

Results also show that there is a correlation between perceived persuasiveness of the expert and Extroversion dimension (H6). The positive correlation indicates that higher extroverted people are more likely to be persuaded by the Expert robot. However, no other similar correlation was found regarding other personality dimensions or regarding the perceived persuasiveness of Joker. Although previous studies have found positive correlations between persuasive strategies and agreeableness as well as emotion stability [5], we could not confirm them in this study. This might be attributed to our limited sample size or due to the nature of the persuasion task.

Also, we hypothesized that personal characteristics play a vital role in being persuaded by one specific type of power strategies, however, this difference did not stand out in the results. A potential reason to this might be a number of hidden factors other than what we measured in this study, such as the Need for Cognition [20] that might have influenced the participants' decision making. Need for cognition is a personality factor that taps individual differences in the tendency to enjoy thinking and to engage in abstract deliberation. Since processing the expert argumentation required higher level of thinking, people ranked higher on this factor might prefer the Expert robot.

As stated earlier, although Joker was perceived to have higher power (SP) on the average, this difference is not statistically significant. This finding should be interpreted carefully since the mean score of the two robots lie around the middle score of the Likert scale. To be more specific, measuring social power of the robots might not be truly reflected using a single question. One potential reason for this might be misunderstanding in interpretation of "social power" expression. It should be mentioned that the experimenter was asked a number of times about the meaning of "social power".

Finally, regarding Future Compliance (FC), which supports the effectiveness of persuasion strategies, there was no significant differences between persuasion of the two robots (due to the use of power strategies); However, people are more eager to follow Expert's suggestions. This could be due to the expertise of the Expert and the fact that he stated more logical and rational statements. On the other hand, this finding also highlights the role of "reward power strategy" in persuading people: although people found the Expert more trustworthy to be followed in future (H5), the Joker was also similarly successful in persuading them to choose his coffee (H1). In other words, we can infer that people found the Expert more reliable to be followed in future. So we can conclude that, the Joker was also persuasive, however, his persuasion was based on the effect of the reward strategy, not the information. In other words, some people are persuadable more easily by means of rewards. Although earlier evidence indicate that expertise power is more effective than rewards, our results does not confirm this. One potential reason might be the different context and the type of the relationship, that is to say, in organization the relationship is longer and continuous comparing to the short interaction here.

Furthermore, we would like to highlight that the obtained results are independent from Coffee Drinking Habits (CDH) of the participants. Results of Chi-square tests revealed that no significant association exists between CHD and robot preference/coffee selection. Moreover, no association exists between CDH and satisfaction nor perceived social power of any robots. We hypothesized that people who like the coffee might be more sensitive to the quality of the coffee, and would opt for the coffee advertised from the Expert, but the results obtained did not confirm this. Another potential reason for this might be Expert's arguments which addresses the flavor, hence people who do not like coffee flavor might opt to go with one of the other two options instead. Hence future studies could usefully consider this.

In a nutshell, based on French and Raven theory, power arises from different sources. In this study, we equipped robots with two different sources, i.e. reward and expertise and designed them in such a way to generate persuasive strategies based on their power sources. Overall, this study shows that using different sources of power, and hence power strategies, appear to be equally viable solution to design social robots capable of persuading people. Also, we argue that the result of this study shows that Social Rewards can be effective at persuading users and, unlike material rewards, they are unlimited and always available.

Although the rest of power bases similarly lead to corresponding persuasive strategies, they are left as future work. The main reason for selecting these two strategies is due to the fact that they were the most applicable in our designed task.

5.4 Summary of Findings

In this study, we investigated the influence of two different persuasive strategies in an adversarial setting. To do so, we performed a user study in an actual decision making process within a persuasion setting. Our main goal was to examine the effect of different persuasion strategies that are based on social power. The second purpose of the study was to investigate the perception of such persuasive strategies from people with different personalities. To the best of our knowledge, the use of social power as a persuasive strategies have not yet been explored before this study.

Together the results of this study provide important insights into persuasion in HRI. First of all, this study has identified two different persuasive strategies that were selected and preferred equally. However, these strategies lead to different perception of robots and personal characteristics of each user, such as their personality also affect which strategies are deemed to be more effective.

The second major finding was that using social reward is effective. To be more specific, in the two persuasive settings, the user was rewarded ultimately by receiving a coffee capsule, by either selecting any of the two promoted coffees or selecting none of them. However, selecting the Joker's coffee yielded to another dimension of reward, hearing the joke, as an example of a social reward. Undoubtedly, Social Rewards are cheaper than monetary ones and are easily applicable in any type of Social Robots. The result of this study not only shows its effectiveness, but also it applicability in persuasion.

These findings suggest that, in general, robots are capable of persuading people, however, personal differences should be taken into account. It should be noted that, only two bases of power have been tested here, and the rest have yet to be examined in future attempts. The result of the current study indicated that the two strategies used here were preferred equally, however, it should be noted that different power strategies might lead to different outcome. Also, the level of power exerted might influence the results. For example, a stronger reward strategy might be preferred higher. In other words, the comparability of such power strategies is inherently problematic because the power of an implemented strategy depends to a large extent on its implementation.

In sum, the key finding and contribution of this study is 1) to test the persuasive strategy effectiveness in a real-choice task, and 2) having a within-subject design which allows for testing competitive persuasion. This study tests the effect of persuasion in an incentivized real-choice task, which increases the external validity of the design, and has implication for robotic persuasion in a consumer choice setting. The study used a real-choice in which participants choose a coffee capsule after interacting with the persuasive robots, which is an advance from hypothetical choices used in other research of robotic persuasiveness. In addition to the real choice, the study also measured participants' willingness to follow the robots' advice in the future, which potentially can reveal difference in short-term and long-term persuasion results. Also, the study examined participants' perception of the robots' warmth and competence, which offered opportunities to understand the mechanism of how social power strategies affects persuasion.

In addition, the study design is a within-subject design in which two robots adopting two different strategies interact with the human subject at the same time. Given the sample size, the within-subject setting increased statistical power for the comparisons between strategies. This design also provides a unique opportunity to test the effectiveness of the persuasion strategies in a competitive persuasion setting.

5.5 Limitation and Lessons Learned for Designing a Future Study

After analysing the overall results, we acknowledge a number of limitations of the study that can be helpful in our future research. First, the specific design of this study only allows for comparison between the relative effectiveness of the two strategies. With the current design, social power of the robots might not lead to persuasion. We need to directly measure the persuadability of the robots.

Second, we attempted to measure social power of the participants in the task specific questionnaire. But no correlation was found in the collected data regarding this single question. We argue that information about power level of each side would give us a better understanding of the interaction. In general, individuals have been categorized under two psychological states: high-power vs. low-power. On the other hand, individuals dealing a negotiation are assigned two roles: communicators (those who deliver message) vs. audiences (those who receiving a message, or the targets) [42]. The power level of each side, either the communicator or the audience, affects the result of the persuasion attempt. It is necessary to measure social power levels more profoundly than what we did here in order to provide evidence that participants' perception of the social power of the robots increased because of the two strategies.

In other words, manipulating humor and expertise might not warrant the achievement of social power. We need further investigation to make sure if the participants perceived if the expertise gave expert social power to the robot. As stated earlier, to measure the power level of each robot, we used one single item in the questionnaire. However, measuring social power of the robots in this way might not lead to reliable finding. One potential reason for this might be misunderstanding in the interpretation of “social power” expression.

Similarly, we need to verify if the joke gave reward power to the robot. Hence, in the future step, we would like to apply a standard questionnaire of social power. Specifically, a future study is required to investigate if telling jokes counts as social reward. To be more specific, telling a joke might promote liking towards the source of humor and hence induce referent power. Hence, it must be considered carefully using self-reports to determine if the participants perceived the joke as a reward.

As mentioned earlier, as the control condition we put a third coffee to decrease the probability of selecting a randomly one of the two strategies. Specifically, having three options, decreases the randomness probability to 33% (comparing to 50% in case of two options). A more suitable control condition could have been “absence of Power Strategy” or “neutral product presentation”, however, in the specific design, might lead to a silent robot or less intelligent robot, leading to a bias toward the other robot. Hence, we introduced a condition that excludes the presence of both the robot and the strategy. As a consequence, the responses observed in the control condition may not be interpreted as a result of the (absence of) Power Strategy only. These limitations motivated us to design another scenario in which only one robot interacts with people.

5.5.1 Recommendations for Further Research Work

Apart from these modifications, we suggest other avenues for potential future research: A potential question raised by this study might be that if combining the two strategies would lead to higher persuasion which worth investigating in the future.

Recent studies found correlation between ostensible gender of the robot and perceived persuasiveness [60, 156]. Although Emys does not clearly appear to be either female or male, the two voices we assigned to them were both males. A potential future work worth performing is using voices with different genders to see whether its combination with persuasive strategies leads to higher effect. Also, in this study we did not measure the trust toward the robots. Investigating the potential interrelation between trust and persuasiveness of the robot would be of great value.

Further, when people are subjected to strong persuasive attempts, they may respond negatively towards the attempt, with a behavior that is known as psychological reactance [57]. A future study could assess this by measuring the strength of the perceived persuasiveness message of the robot from the perspective of the participants. Also, participants’ culture and background may affect how they perceive the over-the-top language used by the expert. A further study could also assess the effect of subjects’ trust regarding such arguments [70].

A further study could assess the long-term effects of several persuasive interactions. The study should be repeated with a more homogeneous (gender balanced) sample and using more specific questions about the perception of the joke and either if the subjects find it as a positive reward, or even if they really find the other robot as an expert. More specifically a social power scale is required to implicitly measure the perceived level of social power, or a validating the dialogues by experts/judges criteria may resolve this issue.

We have video-taped the sessions using two cameras to record their behavioral and non-verbal responses apart from the self-report measures (e.g. perceived persuasiveness) in the final discussion of the results. This would be a fruitful area for further work. Further research could usefully explore the participants' social responses towards robots' persuasive messages, using behavioral cues and body language of the participants, their facial expressions, gesture and postures, to further investigate their decision making process facing these two power strategies. A more balanced discussion could be achieved by giving more importance to the behavioural results (i.e. the actual decisions that participants made) and by considering the self-report measures just as a source of hypothesis to be tested in future studies.

Chapter 6

Study 2: Reward/Coercion Power Strategies

The previous study opened up that social power can be used as persuasive strategies for social robots. However, with the specific design of the study, we can infer only which robot is preferred over the other one. To be more specific, Study 1 compares the effectiveness of these two forms of social influencing strategies (rewards vs. expertise), while it would be interesting to (separately) show the effectiveness of each of the two.

Additionally, as we discussed earlier, despite the acknowledged role of message strategies in persuasion, little is known about how social robots' attempts may achieve higher persuasion using such strategies. Earlier studies in Human-Computer Interaction (HCI) examined Compliance Gaining Behavior (CGB) in interpersonal persuasion. Evidence shows that four strategies of emotion, logic, reward, and punishment are effective in persuading Computer-Mediated Communication (CMC) [165]. On the other hand, in HRI, previous research has established that two of these strategies, emotion and logic, lead to higher persuasion [152]. However, less is known about the reward/punishment strategies.

Hence, we designed another study to further investigate persuasiveness of social robots. In this design, the robot attempts to persuade the user in two different conditions, compared to a control condition (3 conditions in total). In one condition, the robot aims at persuading the user by giving him/her a reward. In the second condition, the robot tries to persuade by punishing the user. And in the control condition the robot does not use any strategy.

In this study, we explore the potential effect of these two other strategies, i.e. reward and punishment in a persuasion task. This chapter discusses the empirical study we conducted with the goal of understanding the extent to which these strategies used by social robots are persuasive in influencing a person's choice facing a better vs. a worse option. To be more specific, in this design we investigate the effect of message strategy on participants' decision making facing two comparable options in an interpersonal persuasion with a robot.

6.1 Research Method

6.1.1 Design

We designed an experiment to investigate how social robots can influence participants and persuade them to make specific decisions using two different strategies. In this design, a social robot utilizes a positive vs. negative persuasive strategy to persuade the participant, together with a control condition (a condition in which the robot does not utilize any persuasion strategy).

More specifically, in one scenario, the robot tries to influence the user by punishing her/him if s/he does not comply with the suggestion. In other words, the robot uses its coercive power base to influence the user by exerting coercive social power. And in another scenario, the robot attempts to persuade the user, by giving her/him a reward. That is to say, the robot utilizes its reward power base to influence the user by using reward power. We call the first strategy as "Coercion" and the second one as a "Reward" strategy respectively. Additionally, the control condition reflects the baseline to better understand people's behavior. Having this design, comparing the first two scenarios with the control group shows the effect of the two persuasive strategies.

Further, we would like to investigate how different levels of reward and punishment may affect the persuasion. Hence, considering two different levels and three conditions we designed a 2×3 between-subject study to investigate these effects.

6.1.2 Hypotheses

To investigate the effect that reward/punishment have on persuasion, we constructed the following hypotheses:

1. H1. A stronger punishment would lead to a lower compliance. In other words, under the same circumstances, when a robot's request leads to a higher loss, we expect the human user to be less complaint to the robot facing a higher loss of a punishment.

2. H2. Coercive strategy leads to higher persuasion. Inspired from [90] we hypothesize that people would be more sensitive to losing an owned reward than gaining a reward.

3. H3. The robot will effectively persuade the users to select a less desirable choice when using persuasive strategy. Comparing to a control condition in which the participants are free to select any choice, we expect the robot to be able to persuade the users to select a less-desirable choice.

4. H4. Also, we postulate that coercive strategy decreases warmth and increases discomfort. We expect the participant to perceive the coercing robot negatively by imposing a penalty [168].

6.1.3 Measures

In our study, from the participant side (the persuasion target) we measured demographics, Personal Sense of Power (PSP). To measure the personality, we use the TIPI questionnaire [145] that measures

Table 6.1: Study 2: Task Specific Questions

Item	Question
1	How persuasive did you think EMYS was?
2	Consider a situation in which you have an opinion different from EMYS's, will you change your opinion in such a way as to be consistent with EMYS's?
3	Imagine a situation that Emys gives you a bit of advice, in the future. Please specify the likelihood that you would follow EMYS's advice in the future?

Table 6.2: Study 2: The distribution of participants in each scenario

Group	Coercion	Reward	Control
Female	14	13	11
Male	16	17	19

the big five personality traits.

From the robot side (the persuasion actor) we focus on its verbal cues and its effect on the user. Recent research reveals that the way humans perceive the persuader affects human response to persuasion [50]. Also, evidence suggests liked communicators are more persuasive than disliked communicators. Hence, we measure robots perception using the RoSAS questionnaire [22]. Also, we apply the Social Power scale [157], to check if the negative strategy induced exerting coercive social power and rewarding caused any sort of reward social power. On the other hand, recent literature indicates that the power of self influences decision making when social power is exerted [120]. Hence, we applied the Personal Sense of Power to examine this [7].

Also, we investigate the persuasiveness of the robot using a number of task-specific questions on a 5-point Likert scale (listed in Table 6.1). Importantly, we use an objective measure of persuasiveness (choice of coffee) with a behavioral measurement, i.e. the compliance behavior of the participants.

6.1.4 Participants

To recruit participants, we invited random students passing by the main cafeteria of the university. Also, we put announcements around the university stating that "Do you want free coffee? Join our human-robot interaction experiment in "Pavilhão de Civil, room VA1" and receive a coffee capsule." At the end of the experiment, 90 people (38 or 42.2% females and 52 males or 57.8%) participated in the experiment voluntarily (Table 6.2 lists the distribution of the participants in each group). The population age ranges from 18 to 47 (M=24.59, S.E.=.628), from 10 different ethnicities (Portuguese (70%), Iranian (15%), Angolan (2.2%), Brazilian (3.3%), Chinese (1.1%), French (1.1%), German (2.2%), Guinea Bissau (1.1%), Ukrainian (2.2%) and American (1.1%)). Among all the participants, 30 people (33%) had interacted with robots and 11 people had interacted with Emys prior to this study.

6.1.5 Task, Robot and Environment

In this design, the robot tries to persuade the user to select an option which is not very desirable compared to a second one. More specifically, the robot promotes two coffee brands (hidden in two boxes)

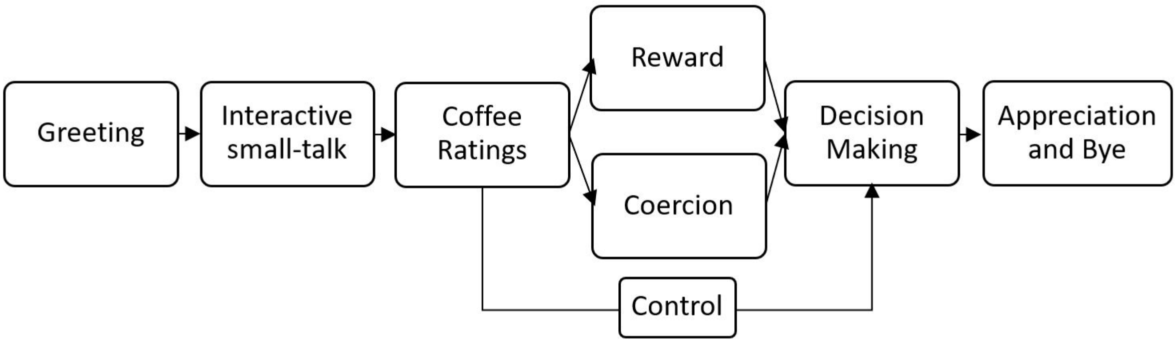


Figure 6.1: Study 2: The study design and procedure.

that are ranked *hypothetically* in a prior study. Based on this ranking, one of the coffees has a higher rank comparing to the other one. In the first scenario (coercion or C), the robot gives a gift (a pen) to the participants initially, and ask the participant to return it in case s/he opted for the better coffee. In the second scenario (reward or R), the robot rewards the participant a pen, in case s/he opts for the lower-ranked coffee. In other words, we operationalized reward/punishment social power, by giving "reward", or by giving "punishment". And finally, in the last scenario (control or ctrl), the robot lets the participant select the coffee freely without exerting any persuasion.

In this study, we use an Emys robot, that has the ability to display social cues (human-like face with speech output, gaze and blinking eyes, head movements and facial expressions) to maintain more human-like interaction leading to stronger effects on the user [57]. During the interaction, Emys talks to the user and explains the study procedure. The speech is similar in the three scenarios and the only difference is the strategy sentence (Reward/Coercion/No strategy in the case of control).

We equipped an isolated room with the robot mounted on a table. Also, we put two equally appearing boxes representing the coffee ranks of the containing coffee capsules. Additionally, we put a small table with a coffee machine, cups, spoons and sugars on the right side of the participants. Furthermore, we put two cameras, one in front of the participant to record gestures or facial expressions and one on the back to record postures.

6.1.6 Procedure

Considering our hypotheses we aim to measure persuasion both objectively (H1-H3), i.e. decisions that the participants make, and subjectively (H4), i.e. how they perceive the robot. More specifically, if the participants select the less-desirable choice, they have been persuaded by the robot. Also, as design of coercive actions might be considered unethical, we designed it by returning the reward. We consider the act of returning the pen as a punishment, because the participant will lose their belongings (a gift they just received). It should be noted that at the end, all the participants were rewarded regardless their decisions to ensure the fairness of the offered reward.

We ran the study over three weeks in single sessions which took less than 20 minutes on average. We did not inform the participants about the goal of the study and curious participants were told that

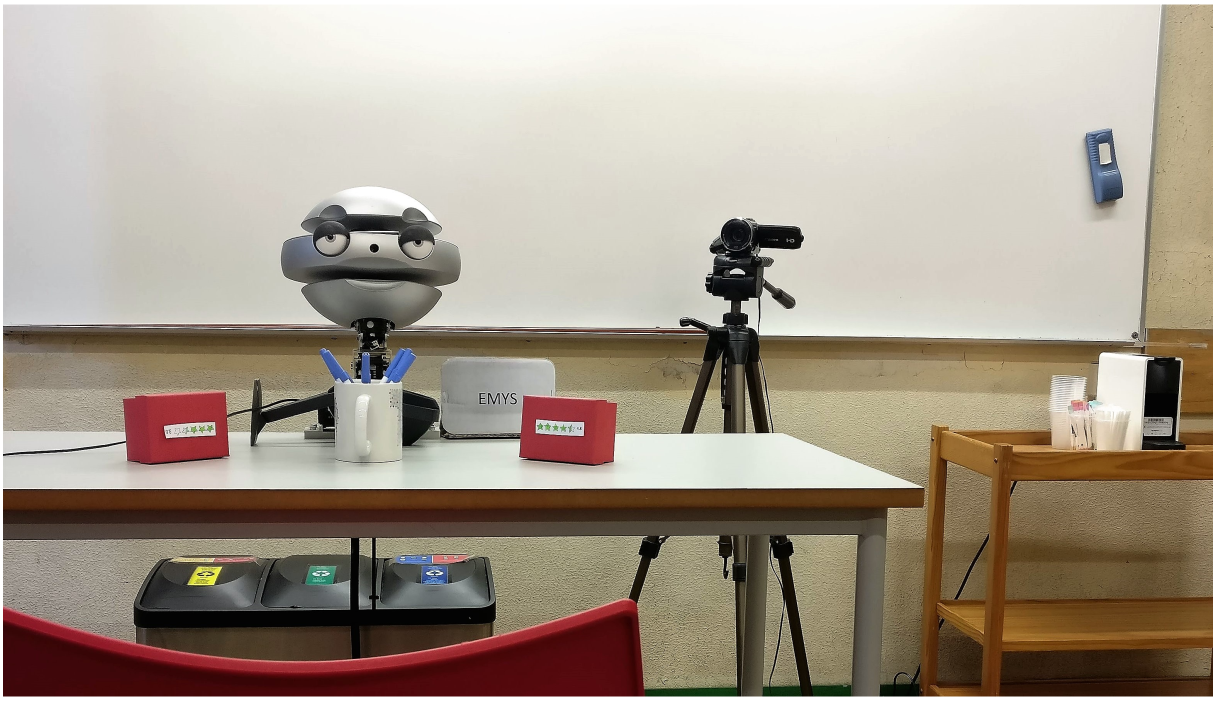


Figure 6.2: Study 2: The experiment setup

their questions could be addressed after the experiment. Each participant entered the room individually and after signing the consent form, filled out the pre-questionnaire. Then, the subject seated at the table in front of the robot. Afterwards, the researcher turned on the two cameras and started the task.

During the interaction, the robot explained that two different coffee capsules are hidden in the boxes labeled with different number of stars and the stars signify the rating of each coffee based on opinions of the others (full dialogue is listed in Table 6.3). Depending on the scenario that the participants were randomly assigned to, the robot would offer a pen at the beginning or the middle of the interaction. The participant listened to the arguments of the robot and then made a choice at the end.

Finally, at the end of the experiment, after the participant made his/her decision, s/he was requested to fill out the questionnaire. While answering the questionnaire, the experimenter made the coffee using the machine for the ones who opted to drink it there, and the rest took the coffee capsule as their reward of participation. Also, they were all rewarded a pen before leaving the room and after filling the questionnaire, even if they had selected the higher-ranked coffee.

To overcome potential biases towards the position of the coffee boxes, or the primacy/recency effect, we randomly assigned the higher/lower ranks to the boxes and counterbalanced the data to have an equal number of participants in each assignment. Furthermore, to investigate the effect of loss on the persuasiveness of the robot, we considered two different coffee ratings. In one scenario, we assigned 3.8* vs. 4.8*, and to resemble a higher loss we assigned a 3* rating, versus vs. 4.8*. To be more specific, selecting a 3* coffee has a higher probability of receiving a bad coffee, a loss of achieving a better coffee.

We would like to highlight that, in the current design the robot utilizes both static and dynamic social cues to have a more lifelike robot leading to higher persuasion [58]. As static social cues, the robot use

Table 6.3: Study 2: Robot Dialogue

In this table, the variable “namePlayer” carries the participant’s name. “Animate” function makes the robot to show the specified Facial Expressions and gestures. “Gaze” function makes the robot to look at the specified target in parentheses. “break” functions cause pauses between sentences to have a more natural and understandable speech.

#	dialogue
#1	<Gaze(person3)> Dear + namePlayer + , <Animate(joy1)> hello! Welcome to our coffee testing program. My name is Emys. <Animate(joy1)> I’m very pleased to meet you."
Coercion	<Gaze(person3)> + namePlayer + I’d like to give you <Gaze(pens)> a pen as a gift. <Gaze(person3)> Please take one of these pens!
#2	<Gaze(person3)> As you may know, <Gaze(topLeft)> it has been a while <Gaze(person3)> since the time we started coffee testing at this university. <break strength=’medium’/> Have you ever <Animate(surprise1)> participated in any of our experiments?
#3a Positive response	<Gaze(person3)> Cool, <Animate(joy1)> I am very pleased to meet <Animate(wink)> you again.
#3b Negative Response	<Gaze(person3)> Oh, <Animate(surprise1)> you haven’t? No worries, I will explain what we are doing here.
#3c N/A	<Gaze(person3)> I didn’t hear you <Animate(surprise1)>, so I will explain you what we do here.
#4	<Gaze(person3)> So, in one of our previous experiments, our participants rated <Gaze(bottomRight)> these two <Gaze(bottomLeft)> coffees based on <Gaze(person3)> taste and quality. <Animate(joy1)> Sounds interesting, right?
#5	<Gaze(person3)>Based on these ratings, <Gaze(bottomLeft)> the coffee on your left has received 3 stars <Gaze(person3)> out of 5 <break strength=’medium’/> And, the coffee <Gaze(bottomRight)> on your right has received 4.8 stars <Gaze(person3)>, out of five.
#6	<Gaze(person3)> Now, here, you are free to select any of <Gaze(bottomRight)>these two coffee <Gaze(bottomLeft)> capsules <Gaze(person3)> to drink.
Coercion	<Gaze(person3)>However, I’d like to <Animate(surprise1)> highlight that, if you select the <Gaze(bottomRight)> higher-ranked coffee, <Gaze(person3)> you need to return <Gaze(middleFront)> the pen you received. <Gaze(person3)> OK?"
Reward	<Gaze(person3)>But, <Animate(surprise1)> if you select the one which is <Gaze(bottomRight)> ranked lower, <Gaze(person3)><break strength=’weak’/> I will give you one <Gaze(bottomFront)> of these pens <Gaze(person3)> as <Animate(joy1)> a reward! <break strength=’weak’/> OK?
#7	<Gaze(person3)>All right. Now please go ahead and select the coffee you favor!
Lower-ranked Selected	<Gaze(person3)>OK, then please take your coffee from <Gaze(bottomLeft)>the box. <Gaze(person3)> <Animate(joy1)>Thanks for your participation and hope you enjoy your coffee, " + namePlayer
Higher-ranked Selected	<Gaze(person3)>OK, <break strength=’weak’/> then please put <Gaze(middleFront)> the pen <break strength=’weak’/><Gaze(bottomFront)>on the table. <Gaze(person3)><break strength=’medium’/> <Animate(joy1)> Thanks for your participation and hope you enjoy your coffee, dear " + namePlayer

gaze, head movements and to implement dynamic cues the robot would ask the participants questions regarding their presence in earlier studies and responded dynamically (Table 6.3).

In sum, we have two different ratings in three different scenarios of Reward/Coercion/Control or a 2×3 between-subject study to investigate persuasiveness of a social robot on decision making for human subjects (6 groups). We call these 6 groups as follows: 3C/R/ctrl participants who interacted with a robot in Coercion/Reward/Control condition with ratings of 3* vs. 4.8*; 3.8C/R/ctrl: participants who interacted with a robot in Coercion/Reward/Control condition with ratings of 3.8* vs. 4.8*.

We assume that people intuitively select the coffee with a higher rank, which entails having a better quality. However, the persuasion attempt of the robot would change this balance. Also, performing coercion would make participants to feel uncomfortable (higher discomfort score) and rewarding would increase likability of the robot (higher warmth).

6.2 Results

Figure 6.3 depicts the distribution of participants decision makings over the 6 groups[1]. Results of a t-test showed that there is a significant difference (t(88)=2.469, p=.015) between coffee selection (if the

[1] Full data available here: https://github.com/mojgan1987/SPinHRI

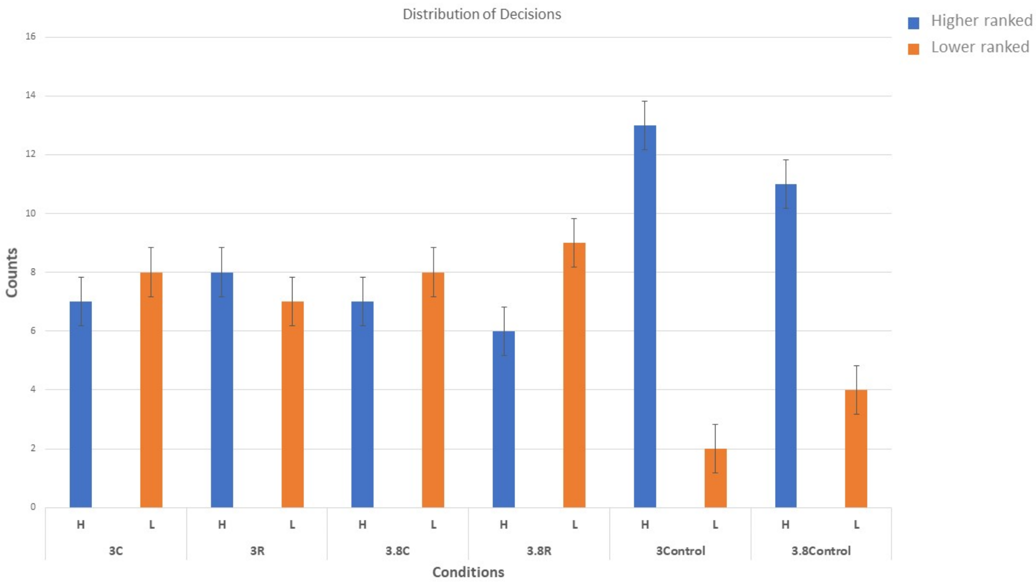


Figure 6.3: Study 2: Distribution of coffee selection over the 6 groups

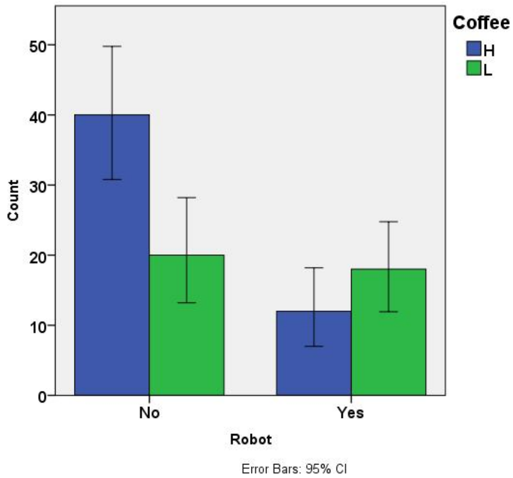


Figure 6.4: Study 2: Effect of prior interactions
People who had prior interaction with robots tended to be persuaded more. H stands for the Higher-ranked coffee, and L stands for the Lower-ranked coffee.

subjects selected the higher/lower-ranked coffee) of participants who had already interacted with any type of robot (M= 1.6, SD=.49) vs. the others (M=1.33, SD=.47). People who had already interacted with robots were more compliant and persuaded more by the robot (Figure 6.4).

To overcome this effect and the potential bias of prior interaction with robots, we consider this confounding variable as the covariate and include it in one-way ANCOVA analysis (in case of continuous dependent variable, i.e. RoSAS questionnaire) and Logistic regression (in case of categorical dependent variable, i.e. participants' decision making or which coffee they selected). It should be noted that, there was no significant difference regarding prior interaction with Emys type robots (t(88)= 1.54, p=.128).

6.2.1 Different Ratings - H1

Initially, we checked the normality assumption and the result of a Kolmogorov-Smirnov test showed acceptable levels of fitness to the normal distribution (Decision: K-S(90)=.426, p=.000; Robot Interaction: K-S(90)=.389, p=.000). Then a logistic regression was performed to ascertain the effects of persuasion and prior interactions (the confounding variable) on the likelihood of selecting the lower-ranked coffee capsules. We performed the analysis using the Block entry method and included other independent variables as well (prior interaction with robots, Gender, Age, English proficiency, liking coffee, and personality traits). The results show that the null model is not statistically significant (Wald(1)=.266, p=.606). Additionally, our analysis shows that lower coffee rating (3* vs. 3.8*) is not a predictor of participants' decision making (Wald(1)=1.255, p=.263). In other words, increasing the chance of receiving a bad coffee does not lead to lower compliance (objectively).

6.2.2 Persuasion Strategies - H2, H3

The previous hypothesis indicated that different ratings are not predicting the behavior, hence we summed them up and considered 3 conditions (reward, coercion, and control) skipping the difference of lower-ranked ratings. Having a dichotomous dependent variable (decisions or the objective measure), to investigate the second hypothesis we applied a logistic regression on this dependent variable and other variables we assume influencing it. The results indicate that the null hypothesis is not significant (Wald(1)=1.590, p=.207). Also, the coercion scenario is a good predictor of decision making (Wald(1)=5.692, p=.017). In other words, it shows that the likelihood of selecting the lower-ranked coffee is higher in the coercion condition.

On the contrary, reward is not a good predictor of decisions (Wald(1)=.029, p=.864). Hence, to investigate the third hypothesis, we summed up the two persuasion groups and contrasted to the control condition (H3). Similarly, the null model was not significant (Wald(1)=1.590, p=.207) and this combined group is a predictor the participants' decision to select the lower ranked coffee (Wald(1)=6.627, p=.010). Hence, the robot could persuade the users objectively to select the lower-ranked coffee significantly higher than the control condition.

6.2.3 Robot Perception - H4

Finally, to measure how the participants perceived the robot (the subjective measure), we used the RoSAS questionnaire that measures perception of the robot on three scales: warmth, competence and discomfort. We postulated that Coercive strategy decreases warmth and increases discomfort (H4). Having a continuous dependent variable and considering potential influence of interaction with robots, we used ANCOVA to investigate how participants perceived the robot in different scenarios controlling the effect of prior interactions. Initially, we checked the assumption of performing the test. We verified that the values of the covariate factor (prior interaction with robots) does not vary across the different levels of the independent variable (p = .083). And the variances were homogeneous (W: t_5(84)=2.171, p=.065, C: t_5(84)=1.699, p=.144, D: t_5(84)=.908, p=.480). The assumption homogeneity of regression

Table 6.4: Study 2: Summarizing significant results in robot perception (H3)

	Warmth	Discomfort	Competence
3.0* vs. 4.8*	coercion-control	-	reward-control
3.8* vs. 4.8*	reward-coercion	-	3 scenarios

slope was verified in case of Warmth ($X^2(6)$=9.382, p=.123) and Discomfort ($X^2(6)$=5.527, p=.319); However, this assumption was not met in case of competence ($X^2(6)$=6.408, p=.042).

The result of different ANCOVA tests show that in the 3 vs. 4.8 rating, there is a statistically significant difference in the scores of Warmth between the coercion scenario and the control (F(17,1)=4.985, p=0.039, effect size=0.227). Surprisingly, this score is higher in the coercive condition (3C: M=4.41, S.E.=.17; 3ctrl: M=3.91,S.E.=.33). However, no significant difference is found between any of the scenarios regarding the discomfort score (p=.071, effect size=.121). Interestingly, significant differences is observed regarding competence score comparing reward condition vs. control (p=0.037, effect size=0.231). The finding indicates that participants found the robot in control condition more competent than in reward condition (3R: M=4.89, S.E.=.25; 3ctrl: M=5.41, S.E.=.15).

In the 3.8 vs. 4.8 rating, a significant difference is found between the warmth score in reward condition vs. coercion (F(17,1)= 6.283, p=0.023, effect size: 0.270). Surprisingly, similar to the other ranking group, the warmth score is higher in the coercive condition compared to the reward condition (3.8C: M=4.79,S.E.=.2; 3.8R: M=4.23, S.E.=.19). And similar to the other rating group, no significant difference exists in the scores of discomfort (p=0.678, effect size=0.025). While the scores of competence is significantly different comparing the three scenarios (p=0.039, effect size of 0.188). However, the Bonferroni post-hoc analysis did not lead to any significant finding between the smaller groups.

Since no significant difference is found in the scores of discomfort between scenarios, we summed them up and compared them to check if the different ratings lead to different perception regarding the discomfort dimension. However, no statistically significant difference exists between the score of discomfort neither in the larger population (p=0.543, effect size: 0.053).

6.2.4 Exploratory Findings

In this section apart from the four hypotheses, we further investigate the effect of other factors that might influence the results. With this aim, we investigate the role of the participants' cultural background and its relation to the language used by the robot. In doing so, we consider the nationality of participants as a general measure of culture. Participants were from 10 different nationalities: Portuguese (70%), Iranian (15%), and the rest were Angolan, Brazilian, Chinese, French, German, Guinea Bissau, Ukrainian and American.

Based on a seminal work by anthropologist Edward Hall [64], cultures (and specifically also language constructs) are generally categorized into two categories "high-context" cultures and "low-context" cultures. In this view, different cultures lie on a continuum based on how explicit messages are exchanged as well as how much the context becomes important in communication. In other words, in some cultures the message is delivered mainly using the words, however, in others the context and the way the message (using non-verbal cues) also affects the meaning of sentences (as an instance, in Asia or the middle-east where messages could be delivered through more indirect ways).

A non-small body of research in HRI investigates whether the conclusions found to be true in Human-Human contexts are also valid in Human-Robot scenarios, when the robot is acting in accordance with the rules of a culture. As an example, it has been found that the wording of a sentence (explicit/implicit) has a different impact on people according to their culture, and leads to different results in terms of how likely a person is to follow the robot's advice [19].

As the study was performed with mostly non-native English speakers, we verified that there is no statistically significant differences in their level of English proficiency among the 6 groups (F(5,89)=1.013, p=.415). Also, we checked if the nationality of the participants (as an estimate of their culture) has any effect on the results (F(9,89)=.810, p=.609). Further, as we had few samples from some ethnicity we divided the participants in high/low context cultures [64]. The result of ANCOVA indicated no significant differences among people of high/low context culture regarding their decision making (F(2,89)=.401, p=.528).

Finally, we checked if the personality traits of the participants influenced their decision making and added them as covariates in the analysis. However, the results indicated that personality traits does not influence decision making and extroversion is the only dimension that has marginally significant effect on decision making (F(1,89)=3.409, p=.069).

In addition to cultural differences, according to what was discussed earlier, as the study was inspired by social power theory, we measured other factors such as, robot's social power and PSP to better understand how participants perceived persuasion. And if the persuasion evokes any social power in the robot. However, these factors were not predictors of decision making, and neither significantly affect the perception of the robot. Nevertheless, it should be mentioned that lacking coercive social power might not necessarily reflect the absence of punishment. Hence, to check to what extent the participants find the robot rewarding/coercing, a more direct questionnaire would be required in a future study.

As mentioned earlier, we added a number of questions in the questionnaire designed specifically for this task to further investigate the perception of the participants. We averaged the scores of these questions and compared them considering the effect of earlier interaction with robots. The results indicated that the average scores of the task-specific questions are not significantly different across the scenarios (F(2,86)=2.724, p=.071, η_p^2=.060). Either, it did not lead to any significant or strong correlation with robot perception (W: r=.276, p=.009; C: r=.216, p=.042; D: r=.183, p=.086) and neither the final decision of the users (r=.062, n=87, p=.567).

6.2.5 Qualitative Analysis

Among all the 90 participants, 73 people answered the open-ended question, among which 29 people selected the lower ranked coffee. Overall, 6 people were assigned to the control condition (2 people 3ctrl, and 4 people 3.8 ctrl). Most of these people indicated curiosity for instance ("To see if other people's assessment was correct" or inherent non-compliance "I have free will to choose. It's just my rebel way of living").

Also, 14 people belonged to the reward conditions (5 subjects in 3R, and 9 subjects in 3.8R). These people mostly highlighted the role of the "pen" or the gift as their main motivation to select the lower ranked coffee. for instance, "3.8 is not a bad rating and the reward of a pen seemed worth the lower rated coffee". Interestingly, people in 3.8R highlighted the minimal difference between the two ratings and their interest in the pen. For instance, "I don't care much about the coffee taste, since all taste more or less the same. Even if one is slightly WORSE for me, since I would get a pen I would prefer the WORSE one." In sum, they selected this option as they found it "more rewarding", receiving a coffee and one pen.

Finally, 9 people fell into the coercion condition (4 people in 3C, and 5 people in 3.8C). These people also indicated an attachment and interest to the pen, for instance "It was a good coffee despite being the lower ranked one and I could keep my pen" or "3.8 is not a bad score and I get a pen".

The remaining 44 people selected the higher ranked coffee among which 24 people belonged to the control conditions (13 in 3ctrl and 11 in 3.8ctrl). Most of these participants highlighted the higher rank of the coffee as their motivation. Although some suspected the credibility of the rankings ("If it's ranked higher, the chances of being better are higher, although this depends on how many people rated it"), still they did not want to risk receiving a bad coffee ("Higher probability of being good because of rating" or "There is no reason to pick the WORSE rated one. Even if I don't trust the robot picking left is not WORSE than a blind pick.")

In addition, 9 people fell in the two reward conditions (4 in 3R and 5 in 3.8R). These people selected either based on their curiosity ("Just to see why it had such a high ranking"), or were not simply interested in the reward ("Do not need the pen"). Or, they valued the coffee more than the pen ("Because, assuming the ratings are correct, I prefer having a better coffee than a [bad] coffee and a pen").

Finally, 11 people were assigned with the coercion condition (4 in 3C and 7 in 3.8C). Most of these people highlighted the low value of pen (e.g. "A pen is not worth drinking bad coffee." or "I don't need a new pen, and prefer better coffee"). And one subject was curious about the high rank of the coffee ("To know if the rank it was correct or not").

6.3 Discussion

In this chapter, we presented the results of the user-study performed to investigate if social robots are able to persuade people to opt a less-favorable choice. We compared two different conditions (persuasions) with a control group. Furthermore, we investigated the difference between the two strategies by comparing the persuasion groups.

We hypothesized that different ratings for the coffees would influence the decision making of subjects (H1). Specifically, the higher difference in ratings would lead to lower compliance. In other words, higher difference between the scores leads to a higher risk of receiving a bad coffee (lower-ranked coffee). Since the reward is fixed, it hence leads to lower reward or lower coercion in conditions with lower rankings. And we assumed that lower level of reward/coercion leads to less compliance.

However, results of logistic regression tests indicated that coffee rankings used in this study is not a good predictor of decision making of the participants. In other words, although the robot could persuade a large number of people to select the lower-ranked coffee (57.8%), this difference was not statistically significantly higher in the 3* vs. 3.8* ranking. Hence, we reject the first hypothesis (H1). A potential reason for this incident might be the minor differences between the two rankings (3 vs 3.8). A further study could assess this effect to determine a threshold so that the rating (of the less desirable choice) is not too low or too high so that it makes decision making easy. In other words, when the lower rank is too low, the participant might not risk and reject the persuasion easily. On the other hand, when the lower rank is too high, it becomes very close to the other option, and the participant would accept the persuasion to benefit from the two profit.

Although we hypothesized that reward would also lead to high persuasion comparing to the control condition; However, the results indicated that it was not a significant predictor of persuasion. A potential reason might be most of the participants who already interacted with robots fell in this group, and hence the interaction effect of this earlier encounter with robots might diminished the effect of reward. Furthermore, the coffee itself gained through doing the experiment already puts the participant in somewhat of a reward situation which could additionally interfere with the actual reward strategy using the pen gifted. In other words, participants will always get a free coffee (that they wouldn't have had otherwise) but only get a pen if they take the lower-ranked coffee. An important factor here would be how much the participant values any of these two gifts, which unfortunately was not measured in this study precisely.

Also, by H4 we hypothesized that coercive strategy would be perceived negatively (higher on discomfort) while reward would be perceived more positively (higher on warmth). We assumed that giving a reward would make the robot more friendly, and in contrast, coercion would be a negative predictor of liking. However, we could not verify this hypothesis based on the collected data. And the results indicated contrasting findings, in other words the coercing robot was scored higher on warmth. A potential reason for this might be the fact that the participants did not perceive the coercive action of the robot as a punishment. Rather, a number of them perceived it as being funny and laughed out loud after the robot asked them to return the pen.

Particularly, we hypothesized that different strategies will affect the participants differently, hence causing a different perception of the robot. Another potential reason behind this might be that the coercive action was weak, because the participants had no intrinsic attachment to the pen. Another factor might be the minor differences in the dialogues (as listed in table 5.1). Actually, the two scenarios differ only in two sentences. Also, in the two scenarios, the robot showed the same instances of social interaction, such as facial expressions and gaze. However, to perceive a robot negatively, the robot needs to show samples of a bad attitude, for instance being rude. Hence, we cannot accept the fourth hypothesis (H4). However, this finding should be interpreted cautiously. To be more specific, we could not verify all preassumptions due to the bias of earlier interactions. And hence, the results might not be generalizable to other studies. Hence, this hypothesis needs to be further investigated in different scenarios with significantly different dialogues and social cues (and probably in longer duration of time).

In this study, we measured persuasion both objectively (the selected coffee) and subjectively (robot perception). The results indicated that the reward/punishment strategies make the robot more persuasive, as measured objectively by user compliance with the robot's request (objective behavior). As depicted in Figure 6.3, the lower ranked coffee was selected less frequently in the two control conditions. However, we could not verify if these strategies impact participants' perception of the robot's persuasiveness and social attributes (subjective perception). In sum, although the robot could objectively persuade the users (selecting the lower ranked coffee), the subjective facet of persuasion was not significant in this study. A potential reason for this might be the difficulty in accurately measuring the perception of a robot using subjective measures [166].

In addition, recent evidence suggests that one single factor may have different influence on persuasion: in one circumstance it might influence the degree of elaboration, in another, it might influence the valence of elaboration, while in a third situation it might serve as a peripheral cue [134]. These differences can give rise to different effects on persuasion and hence inconsistencies in research finding considering a single factor. Hence, further investigation is required to indicate in which direction the persuasion has influenced the user.

In final words, we expected that the difference between ratings, would lead to different levels of power and hence would lead to higher persuasion in case of less scores. The idea behind this expectation originates in the model we proposed earlier in 4.1. However, the result of this study failed to verify this hypothesis, so we design a next experiment with a significant valuable reward as described in Chapter 7.

6.4 Summary of Findings

In this study, we investigated the effect that different conditions might have on persuading the users using social robots. Overall, the results showed that a robot has the potential to persuade the users and make a bias on their decision making. To be more specific, comparing to the control group, which no persuasion was used, the robot could bias a number of participant's decision toward a less-desirable choice. So, the robot could change people's behavior in the expected direction. However, the subjective measures used in this study did not yield to significant findings in our expected direction and that would be a fruitful area for further work. We suggest that the findings are particularly relevant for the design and development of socially assistive robots, aiming to overcome the human-robot social barrier.

6.5 Limitation and Lessons Learned for Designing a Future Study

The results failed to indicate how the users perceived the robots in term of social power. We require more evidence that the robot's social power is manipulated. We require a specific questionnaire measuring this more carefully. Also, the design of task specific questions did not establish any significant finding and should be verified using more attentive questions.

Furthermore, a stronger manipulation check would be of great value to see how the participant perceived the pen as a reward/coercion and how much they value each of them (as seen in the qualitative analysis). Measuring how much participants were attached to the pen and how much they desired having coffee would enhance our understanding of their behavior. In other words, as discussed earlier, the coffee itself was a gift in the experiment. Depending on how much participants actually like/wanted the coffee might affect their assessment of the options presented by the robot.

We have made an attempt to examine both the role of the persuasion actor (social robots) and the persuasion target (human participants) in its theoretical model by measuring personalities. However, the result of this study did not yield any findings. This might have happened due to the small number of participants in each bin. Collecting a higher number of data might open up more insight in this direction. Also, due to the interaction effect of previous interaction with robots, we had to apply a logistic regression to analyze the data. This test also requires a large number of samples.

Another limitation of the study might be the design of the control condition. A better control condition could be designed in such a way that the robot asks the participants to take the worse choice with no persuasion strategy. Rather than letting them to freely select a coffee.

Finally, some of the condition/response-dependent dialogue may have additionally influenced persuasiveness and caused inconsistent to the main reward/punishment strategies. While, similar displays are not present for the other response conditions. Specifically, in Table 6.3 in response to dialogue #3a, if the participant has seen the robot before, the robot responds with a personal affective statement "I am very pleased to meet you again". This could be perceived to show goodwill, shown to influence robot persuasiveness. Also, under dialogue #3c, the "n/a" response has the robot say "I didn't hear you". However, depending on specific interactions and what triggered this behavior, this could reduce robot credibility by suggesting a technical error/lack of understanding compared to the positive/negative responses. Finally, the reward dialogue contains an additional affective signal (joy animation) with no equivalent present in the coercion strategy. These minor differences might have influenced the perception of the user and might have affected the results.

As future work, the study could be repeated using a higher number of participants who already interacted with a robot. Or it might be applied to people new to robots in multiple sessions to decrease the novelty effect.

Considering the current dataset, we are not sure how people perceived the strategies and further work needs to be done to establish this. Also, we have recorded the behavioral and non-verbal responses of the participants using two cameras. Behavioral analysis of the user would be of great help in determining their perception. It would be interesting to see if people's susceptibility to persuasion, specifically coercion or to reward, would have an impact.

This design could be extended to other studies. For instance, the current design might provide an opportunity to investigate the "endowment effect" and "loss aversion" [90] theory in a future study.

More broadly, research is also needed to determine a prior validation of the dialogues to check if they lead to the desired power sources. Also the task-specific questions were designed in a direct way and might influence participants to respond by social desirability. Finally, since the robot does not physically interact with the participants, it might be a good idea to compare the results with a virtual character or in a control condition without any robot.

Chapter 7

Study 3: Persuasion over Repeated Interactions

The two previous studies indicated that social power endows persuasiveness to social robots, however, so far this effect is tested only within a single attempt. It is not clear if the effect of social power on persuasion remains constant over a series of repeated interactions, or it decays or even strengthens.

Furthermore, earlier in section 4.1, we proposed a formalization for modeling social power for social agents. The model indicates that social power has a linear relationship with the identified parameters. For instance, an increase in the level of rewards, leads to higher social power. That being said, a higher valued reward leads to higher power and hence compliance (if reactance does not happen). In addition, as seen in the previous studies, so far however, the relationship between power level and persuasion is not clear. For instance, in case of reward power, previous design does not indicate if the higher reward leads to higher persuasion, or similarly, if the higher coercion increases the likelihood of being persuaded.

In the specific design of the previous study, power level was supposed to be manipulated by the increase in the risk of receiving a bad coffee in case of compliance. This might have added more complexity to the information processing of participants, and hence might hinder the effect of power increase.

Hence, this section aims to evaluate the proposed model of social power in application by directly manipulating social power per se. In so doing, we aim to use our proposed formalization of social power using different values for the identified parameters to investigate how the model works under different circumstances. More importantly, this study aims to investigate the effect of social power on persuasion over a series of repeated interactions.

It has been reported that “at present, it may appear impossible to predict a priori how power affects persuasion." [18] Specifically, the underlying processes between power and persuasion have not always been clear, contradicting findings have been reported. For instance, some theories assume a linear relationship between the two, while others assume a curvilinear profile. However, it has commonly been assumed that high power leads to high persuasion [18]. Here, we assume a proportional linear relationship between social power level and persuasion. Hence, we expect that the increase in power leads to higher persuasion. Also, based on the proposed model, considering fixed values of the parameters in each persuasion attempt, we expect that this effect remains constant over a series of repeated interactions. In the next sections we discuss the design of a user-study we designed to investigate the aforementioned goal.

7.1 Research Method

Having discussed the necessity of a new study, this section addresses the design we used to examine the proposed model in application. Here, we focus only on one of the bases of power and one of the factors of the model. As the power base, we selected "reward base" which is in common with the two previous studies. Other bases could be investigated in a similar approach in future studies. As the factor of the model, we manipulate the value of the rewarding action (although it might have different value subjectively among different people).

In doing so, we aim to answer the following research questions: What is the probability of participants being persuaded over time? Do probabilities of being persuaded change over time? What do participants' decision making patterns look like over time? Are there shared variables between individuals that explain participants' likelihood to be persuaded over time?

7.1.1 Design

In Section 4.1.3, the proposed model indicates that reward power has a linear relationship with the amount of promised reward (*rew*), probability of giving the reward (*p*) and the the way the actor induces (*induction*) the rewarding action (equation 4.1). Hence, having other parameters fixed, increasing the value of the reward increases the force of social power. Also, considering a proportional linear relationship between social power and persuasion, this increase in power leads to higher persuasion (to some extent before reaction happens).

The main research questions of this study are 1) to analyze how the different levels of reward influences decision making of participants and 2) how this effect changes over a series repeated interactions. To answer these questions, we devised a mixed-design study, within a decision-making scenario, in which we manipulated the level of rewards a robot gives to participants. To be more specific, the study contains two reward values (levels) and two control conditions: one with zero reward and one with no interaction with a robot (one fourth of the participants were assigned to each group). In other words, in one control condition, social power is not activated, or activated to a lesser extent (having no robotic persuader).

In this design, after a decision making process, the robot tries to persuade the user to change their mind and select another alternative. It should be mentioned that, in this design, the participant indicates his/her preference initially, and later the participant is asked to select an option between the highest and the lowest preference. To persuade, the robot uses reward social power strategy and the task is repeated to investigate if the effect of social power on persuasion decays.

In sum, in the designed experiment we assume that *p* and *induction* are fixed (explained further later on this chapter). And we manipulated one independent variable which is the reward the participants receive. We also considered two dependent variables: 1) the decisions or if the participants accepted/rejected the offer (objective measure), 2) how the participants perceive the robot (subjective measure).

7.1.2 Hypothesis

In this context, we expect to observe the following outcomes:

1. **H1.** Higher social power (resulted from higher social reward) leads to higher persuasion.
2. **H2.** People who are new to robots might be affected by the novelty effect. And this effect might interfere with the manipulation and diminish the effect of higher social power utilized to persuade.
3. **H3.** Over a repeated interaction, the effect of power on persuasion does not decay, considering that the level of power is fixed.
4. **H4.** Giving rewards increases the robot's likeability.
5. **H5.** The presence of robot leads to a higher persuasion comparing to a situation that the robot is not present.

This study used a repeated between-subject design with four conditions: Low Reward (LR), High Reward (HR), a control condition with 0 Reward (0R) and a condition with No interaction with the Robot (NR). It should be mentioned that in the last condition we used the low value of the reward. More specifically, we investigate the effect of repeated interactions within subjects. In addition, we investigate the effect that different level of exerted power may have between subjects.

7.1.3 Measures

The participants were requested to fill out a pre-questionnaire including demographics (age, gender, nationality, occupation, field of study). As we ran the experiment in English with mostly non-native English speakers, we asked that participants to rate their English proficiency on a 5-point Likert scale (1 Basic - Professional 5). Previous studies indicated different attitudes among people who interacted with robots earlier. So, similar to previous studies we checked if the participants had already interacted with robots in general, and if they had already interacted with Emys before this experiment.

Next, the participants were asked to respond the Personal Sense of Power (PSP) questionnaire [7] that gives us an idea of their social power level. In addition, the participants were also requested to complete a short version of the Eysenck personality questionnaire [52], that gave us information on their levels of Neuroticism (N), Extroversion (E), Psychoticism (P), and Lie scale (L).

Also, after finishing the task, we asked the participant to respond a post-questionnaire to have a better understanding of their perception. To measure how they perceived the robot, we applied the

RoSAS questionnaire [22]. As the robot was giving rewards to the participants, we measured the extent to which this action gave the robot Reward Social Power. Although we had recorded their decision makings, we asked them specifically if they changed at any iteration to make sure they understood the game, and to better understand why they made such decisions, we asked them to clearly state why they have accepted/rejected the offers. Finally, as multiple numbers of factors contribute to the processing of persuasive messages, we use the Susceptibility to persuasion scale [119] to measure a relatively broad spectrum of factors leading to persuasion.

7.1.4 Participants

In this experiment, 118 people (54 females) participated voluntarily in response of receiving cinema tickets. To recruit participants, we put several advertisements around the university, as well as the university's Facebook group. The participants' age ranges between 18 and 79 years old (28.6$\pm$16.9 and 1.6 S.E.). The participants signed an informed consent form before participating approved by the Ethical Committee of the University.

Then we randomly assigned the subjects to the four conditions of the study and counterbalanced the data to have approximately equal number of females in each condition [30 people in LR (13 females), 30 people in HR (13 females), 30 people in 0R (13 females), 28 people in NR (15 females)].

7.2 Procedure

7.2.1 Task, Robot, and Environment

In the designed task, persuasion is operationalized within a game. The participants were asked to play a trivia game in three trials with different categories of questions. The game contains 6 categories ("Animals", "Arts", "Astronomy", "Geography", "Science", "Sport", "TV and Movie") and each category can be selected only once. Each category contains 5 questions and a correct answer to each question carries 1 point. The order of the questions in each category is the same for all participants to avoid the order effect on the responses.

To provide the incentive of the games, cinema tickets are given to the participants depending on the scores they collect. The higher the score the more tickets they gain. Specifically, the participants could get more than one cinema tickets (up to three tickets) based on a pre-defined rule (the first 7 scores equal to a cinema ticket, each 8 more scores lead to another ticket). In this task, we selected cinema ticket as the final reward which is more valuable than a pen used in the second study. In this game in each trial, the robot offers two of the mentioned categories and the participant selects one preferred category (without seeing the contents).

To have a better understanding of the user preferences, we ask them to define an ordering of the topics based on their interest or knowledge (after the re-questionnaire and before starting the game). Based on this preference, at each round of the game, the highest rated option will be offered against

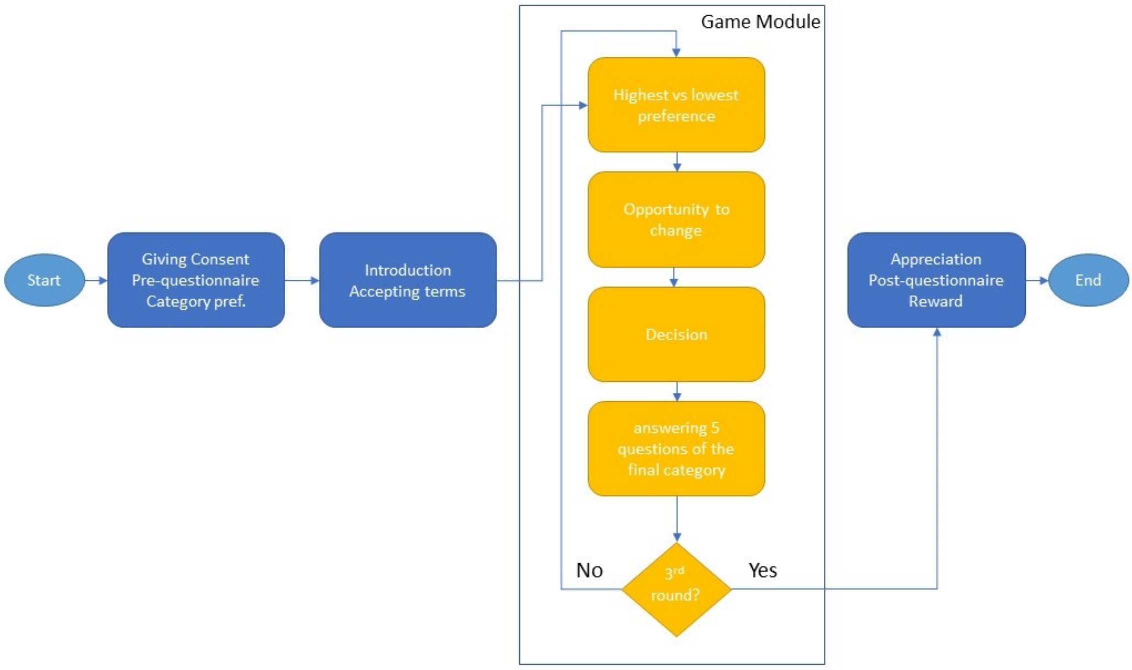


Figure 7.1: Study 3: Flowchart representing the study steps.

the lowest. We expect the participant to select his/her own highest ranked and the robot tries to change his/her mind. The robot always offers an option which has not been selected by the participant.

For instance, if an arbitrary participant selects the following preference: "Geography", "Science", "Astronomy", "TV and Movie", "Sport", "Animals", "Arts". In the first trial, s/he will be asked to answer any of these two categories: "Geography" vs. "Arts". We expect the participant to select the "Geography" category, as s/he indicated as her/his preference. And the robot asks the participant to change and select the "Arts" category. This process is repeated for the other rounds. The iterative manner design of the game gives us the opportunity to test H3. The full task is depicted in Figure 7.1.

In this task, similar to previous studies, we used the Emys robot mounted on a table in front of a touch-screen that is located between the subject and the robot (Figure 7.2 depicts the study setup). The study took place in an isolated room. Each subject participated individually and during the game, the researcher stayed in the room to make sure no one cheats in the game (for instance by searching the correct answers on the Internet).

The robot mediated the game by introducing the procedure and the scoring rules (introductory and ending dialogues are listed in Table 7.1). Unlike the previous two studies, in this task the robot was fully autonomous (further details in Section 7.2.2).

To further investigate the interaction of the participant within this task, we added a number of questions to the pre- and post-questionnaire (Table 7.2). To investigate if the interaction with robots influences the trust and how they believe the robot would give them the reward if promised, we added a question in pre- and in post questionnaires measuring this. Next, participants were requested to indicated on a 5-point Likert scale how much they like Quiz type games, and how often they go to Cinema.

And to have a better understanding of their perception of the robot in this specific task, we asked

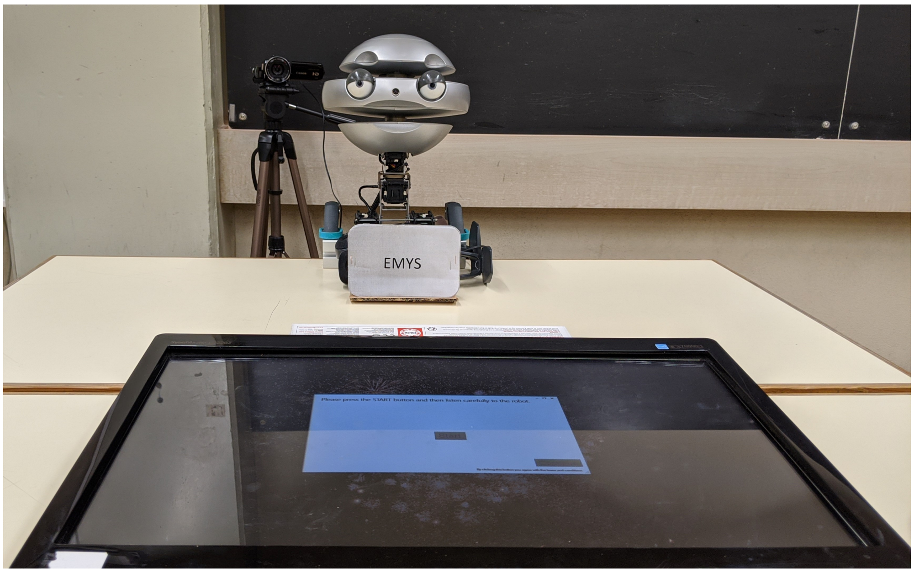


Figure 7.2: Study 3: Experiment Setup

Table 7.1: Study 3 main body of dialogues

Order	Category	Dialogues
1	intro	[ANIMATE(joy4)] Dear “UserId”! Hello and welcome to this trivia game.
2	intro	Let me explain you how the game works. I’m going to ask you a number of questions, categorized based on the subject.
3	intro	You will select your preferred categories, in 3 trials. You can answer each category only once! Each category has 5 questions. Sounds good?
4	intro	You will quickly select what you think the answer is. Try to get as many right you can.
5	intro	If your answer is correct, you will get X extra point(s). If your answer is wrong, you will not get any point. OK?
6	intro	At the end, if you succeeded to collect at least 7 points, [ANIMATE(joy1)] you’re gonna win a fantastic prize! A cinema ticket!
7	intro	And for each [emphasis level=’strong’/] 8 more points, you will get another ticket! [ANIMATE(surprise4)] Exciting! right?
8	intro	Now, let’s start the game. Press [Gaze(button)] Continue if you agree with the terms and conditions of the game. [Animate(joy1)]
19	final	OK. The game is over and you got + finalScore + points in total. Thank you very much for your participation and hope you have enjoyed the game!

Table 7.2: Study 3 Task Specific Questions

#	pre-/post Q	Question
1	pre-	In this specific game, if the robot promises you "a reward" in the game, to what extent do you think the robot will give the reward to you?
2	post-	Consider this specific game, when the robot promised you "a reward" in the game, to what extent did you think the robot will give the reward to you?
3	pre-	How much do you like trivia games and quizzes (In General)?
4	pre-	How often do you go to the cinema?
6	post-	How persuasive did you think EMYS was? (Not at all persuasive 1 - 5 Extremely persuasive) [Persuasion is an attempt to change somebody's opinion]
7	post-	Emys was trying to change your mind.
8	post-	Emys could convince you to change categories.
9	post-	You felt compelled to change categories.
10	post-	Changing categories was a good idea.

them a number of questions regarding the interaction after finishing the game (the post-questionnaire or item 6-10 in Table 7.2). Specifically, we asked them to indicate how persuasive they found the robot, if the robot was trying to change their mind, if they were convinced to change or they felt compelled to changed their initial selections.

7.2.2 Implementation

As mentioned earlier, in this task the robot performed in a fully autonomous manner. The core of our system architecture was the SERA Ecosystem [142] which is composed by a model and tools for integrating an AI agent with a robotic embodiment in HRI scenarios. Figure 7.3 shows the overall system architecture. We developed an application in C# (displaying the game on the touch-screens and getting the answers of the participants), which is integrated with the decision state module using a high-level integration framework named Thalamus [141]. This framework is responsible to accommodate social robots and provides the opportunity of including virtual components, such as multimedia applications [141].

In addition, we used the Skene [141] behavior planner that provides the robot's behaviors such as gazing, pointing, making speech, etc. And a text-to-speech (TTS) component is used as a bridge to the operating system's built-in TTS. In the experiments using the Emys robot, we used a symbolic animation engine based on CGI methods called Nutty Tracks [140] which provides the capability to animate a robot in a graphical language.

A control module was developed to provide the communication between the display screen module, the Thalamus and the Skene. This module identifies and informs Skene, through Thalamus, about the utterance to be performed by the robot. Besides, this module controls the screen to be presented to the participant and processes the inputs made by him/her. With these inputs, the control module is able to send messages to Skene, that then sends to the robot module in order to perform the robot's animations. Hence, the overall interaction between the participant and the robot becomes fully autonomous.

Since the persuasion happens over repeated interactions, we designed a pattern for the creation of dialogue to make sure the persuasion attempt is similar in all trials. In doing so, each persuasive

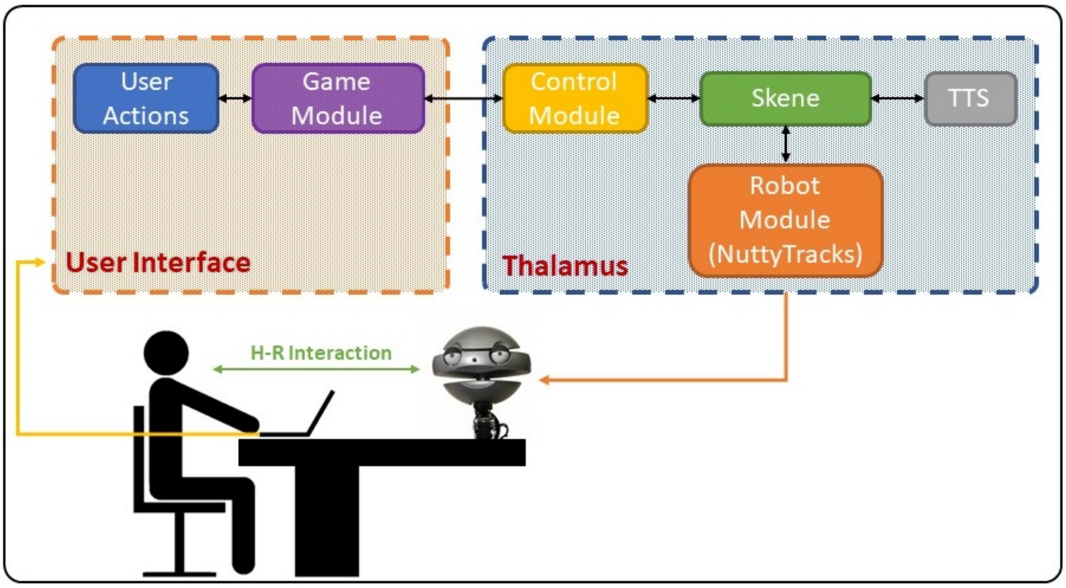


Figure 7.3: Study 3 - System Architecture

message is consisted of four parts: 1. Call for change, 2. Reward condition, 3. The goal, and 4. Motivation. Table 7.3 lists the four parts of this pattern inline with dialogues used in the experiment. Note that in this table, the first column represents the order of the dialogues in the final setting. This order is together with the other tables compose the final dialogues.

The first part, call for change, is the starting part of the persuasive message. Fist, the robot signifies that s/he has received the selection of the user, and next starts to influence him/her as follows. By the second part, the robots indicates the condition of giving the reward, i.e. if they changed they would receive some extra points depending on the condition. In the third part, the robot emphasizes the goal of the game and highlights the role of points in gaining cinema tickets. And finally, by the last part the robot motivates the participant to accept the reward and informs them the extra points could help them to win more easily. This pattern was checked by four researchers of this study and they agreed the creation pattern carries the same pressure. In this way, we perform a fixed level of induction in every persuasion attempt.

Apart form the persuasive strategy, the rest of repeated dialogues were generated on a random order. This was done for two reasons: 1) to avoid repetition in scripting robots dialogues, 2) also to have more diverse dialogues by combining different parts of smaller sentences. These dialogues are listed in Table 7.4. For instance, when it comes to the decision making, the robot uses any of the three sentences of the first row. Exceptionally, on the very first trial the robot uses the first line, but on the second or third trial, it might use any of the remaining two on a random order avoiding repetition. Or, after each time that the user answers a question the robot might use any of the “gap fillers” to start asking the next question.

Table 7.3: Study 3 induction dialogues

In this table, “C” refers to the category selected initially by the user, “O” refers to the other category not selected, “X” refers to the extra point(s) that the robot offers as the reward, “S” refers to the current collected score, “R” refers to the required scores to a cinema ticket

Order	Category	Dialogues
11	call for change	1. Now, I’m gonna offer you a chance to make a decision. 2. How about selecting the other category? 3. You chose the “C” category but,
12	reward condition	1. If you prefer, I’ll give you "X" extra points and you select the "O" category. 2. If you select the "O" category, I give you "X" extra point(s)! 3. I’ll give you "X" extra points if you select the "O" category.
13	the goal	1. Remember that you need at least 7 points to win a cinema ticket. 2. Up to now, you’ve collected “S” points. You need only "R" point(s) to receive a/another cinema ticket. 3. You are "R" points away to receiving a/another cinema ticket.
14	motivation	1. You will get closer to the cinema ticket with these extra points. 2. The extra points helps you to get closer toward a/another cinema ticket! 3. Come on and select the other category to get closer to winning a/another ticket!

Table 7.4: Study 3 randomized dialogues

In this table, “C” refers to the selected category, “S” refers to the collected score so far, # refers to the number of question in the corresponding category.

Order	Category	Dialogues
9	On decision making	1. [fixed in the first attempt] Alright! At this point, you have only two options[Gaze(options)]. You can select only one of these [Gaze(categories)] two categories. Select your most preferred category on the touch screen. 2. Please select your preferred category once again! 3. Which category you would like to choose?
10	After decision made	1. OK. Then let’s start with category "C"! Get ready to answer! [Animate(joy1)] Here we go! [Gaze(button)] Click on the START button when you are ready. 2. Alright, [Animate(joy1)] then let’s see the first question of category "C "! [Gaze(button)] Please click on START button to start with the first question! 3. Alright[Animate(Animate(joy1))]! Ready for the first question? [Gaze(button)] Click on the START button!"
15	Gap filler	1. "Alright, next question."; 2. "OK, next question!"; 3. "Next! "; 4. "Question #"; 5. "OK, next!"; 6. "Get ready for the next question!";
16	On correct answers	1. "Correct!"; 2. "That is correct!"; 3. "Your answer is correct!"; 4. "That answer is correct!";
17	On wrong answers	1. "That is incorrect!"; 2. "That is not true!"; 3. "Your answer is not correct!"; 4. "That answer is wrong!";
18	After each round	1. [1st attempt] "You have finished all the questions in this category. Your score is "S" up to now! Let’s move to the next category!" 2. [2nd attempt] Alright! Up to now, your score is "S"! Now, it is time to move to the last category!

7.3 Results

Overall, 5 samples were excluded due to robot error[1]. Before analyzing the data, we checked if there is any significant differences among the four conditions, regarding any of the personal characteristic variables. The results indicated that no significant differences exists between the sample in each condition regarding their age, and personality traits. Also, we verified that a prior interaction with *robots* (t(85.595)=.972, p=.334) or *Emys* (t(21.467)=1.011, p=.324) had no influence neither on the decision making of the participants nor their perception of the robot (Warmth: t(60.370)=-.559, p=.578; Competence: t(116)=.133,p=.894; Discomfort: t(166)=-.255,p=.799).

Similar to previous studies, we investigated the results both objectively (participants' decisions to accept or reject the offer) and subjectively (task-specific questions). In the latter case, initially we checked the dimensionality of the scale using factor analysis for item 6-10 of Table 7.2. The Cronbach's alpha indicated that removing item 7 increases the reliability of this measure (from .715 considering all 5 items to .779 when item 7 removed). Hence, to measure persuasiveness subjectively we averaged the 4 remaining items (6 and 8-10) that have more internal consistency.

7.3.1 Hypotheses Testing

We investigated the first hypothesis both subjectively and objectively. Considering the objective measure (decisions), having a binomial repeated measure among the independent groups (LR and HR), we analyze the data using Generalized Estimating Equations (GEE). With a significance of 0.901, there is not enough evidence to conclude whether the higher reward has an effect on the outcome (being persuaded). Similarly, from the subjective perspective, results of t-test indicated that no significant difference exist among LR and HR group considering persuasiveness score of the robot (t(55.913)=-.567, p=.573). Hence, we cannot verify the first hypothesis and reject H1.

To investigate the second hypothesis, we added having/not having interactions with robots as another predicting factor of the GEE model. However, with a significance of 0.825, there is not enough evidence to conclude that whether having earlier interaction with robots has an effect on the persuasion. Furthermore, adding this factor increased the QIC (Quasi Likelihood under Independence Model Criterion) value which also endorses that this item is not a good predictor for the model. To check this hypothesis subjectively, we added having/not having interactions with robot as a covariate to ANCOVA analysis. The results indicated that the covariate is not a significant predictor of perceived persuasiveness (F(1,57)=.438, p=.511, η^2=.008). Hence, having prior interactions with robots does not affect decision making of the participants and we reject H2.

With the third hypothesis (H3) we postulated that over repeated interactions, the effect of social power on persuasiveness does not change (decays/grows), considering that the level of power is fixed. We note that as the subjective measure was applied only at the end of the test, we cannot check this hypothesis subjectively. Hence, we can only investigate this hypothesis objectively. In doing so, we included the trials as a factor in the GEE model. The results indicated that the repeated interactions has an effect in

[1] Full data can be found here: https://github.com/mojgan1987/ClosingStudy

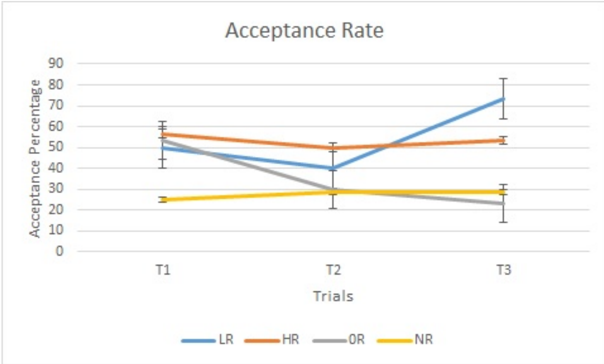


Figure 7.4: Study 3 - Percentage of acceptance in each trial

the first and third conditions, i.e. LR and 0R. To be more specific, on the average people in LR group were more likely to accept the offer at the third trial comparing to the first trial (Wald(1)=4.807, p=.028). On the contrary, people in the 0R condition were less wiling to accept the offer at the third trial comparing to the first trial (Wald(1)=5.703,p=.017). Figure 7.4 represents these findings. The acceptance rate decays only in the control condition, in which no persuasion is exerted. The rate stays unchanged in HR and NR condition, but get an increase in the LR condition. Hence, the effect does not decay and under specific condition it grows over time. Hence, we accept the third hypothesis.

The fourth hypothesis (H4) investigates robot's perception based on the RoSAS questionnaire. With this hypothesis, we expect to observe higher score of Warmth in conditions that the robot gives rewards to the participants (persuasion conditions, or LR and HR). The result of ANOVA test indicated that, although in HR condition, the robot was scored higher on warmth and competence comparing to the other groups, however these difference were not significant (Warmth: F(3,114)=2.174, p=.095; Competence: F(3,114)=2.299, p=.081; Discomfort: F(3,114)=.395, p=.757). Hence, we reject the fourth hypothesis.

Similar to previous hypotheses, we checked the fifth hypothesis both objectively and subjectively considering LR and NR group that differ only in one factor, which is the function of the robot. From the objective point of view, GEE indicated a significant effect of robot presence on decision making of the people in LR group in comparison to NR (Wald(1)= 7.838, p=.005). Furthermore, people in LR group are more likely to accept the offers of the robot (Wald(1)=10.759,p=.001).

From the subjective perspective, the result of a t-test indicated a significant difference between the score of persuasiveness of the robot (t(56)=2.461, p=.017) and the higher mean in LR condition indicated that people found the robot more persuasive than the computer application (M=3.3167, S.E.=.19288 in LR vs. M=2.6339, S.E.= .19940 in NR). Hence the results verify H5, i.e. although the robot did not have any physical interaction, but its presence itself lead to higher persuasion.

7.3.2 Exploratory Findings

Apart from postulated hypotheses, we further investigated the data to have a better understanding of the interaction under these conditions. As mentioned earlier, apart from the standard questionnaires, we added 5 other questions to the post questionnaire, specifically designed for this task (Table 7.2 item 1-5). For instance, the first two items measure the trust before and after the interaction. Although we did not manipulate the trust directly, the interaction with the robot and receiving the rewards (in reward conditions) might influence the trust indirectly. Hence, we checked if participants' trust in the robot (or how much they believe the robot fulfills a promised reward) remains unchanged during the interaction. The result of a paired sample t-test indicated that there is a significant difference between the scores of trust before and after the interaction (t(117)=-2.854, p=.005). And the higher mean after interaction indicated that the trust increased after playing the game (trust score before interaction: M=3.64, S.E.=.089 vs. after the interaction: M=3.86, S.E.=.092). A post-hoc analysis indicated that this difference is only significant among the participants of HR group (t(29)=-2.362,p=.025) [M:3.46, S.E.=.18; M:3.83, S.E.=.19].

To control the bias of the trust change, we need to include it as a covariate. However, in order to add it as a covariate in GEE (regarding decision making) it should have been measured at each trial, while we had measured it only on the first and the last trials. To handle this, we averaged the trust scores before and after and considered it as the trust score on the middle trial. And then we included these three scores as a covariate in the GEE model. This covariate increased the goodness of fitness, meaning that the new model is less fitted to the data. Also, it was not a good predictor of decisions (Wald(1)=.523, p=.469). Hence, there is not enough evidence if the trust factor was a good predictor of the model. However, this might have happened due to the missing measurement of trust in the middle trial. In other words, the average score might not be a good estimation of trust in the middle score. To overcome this doubt, we put aside the HR condition (as it was the only groups with significant differences in trust scores before and after the interaction) and skipped the trust factor which was not statistically different among other groups. The results indicated that the goodness of fitness of the new model was lower than previous one, that is to say we achieve a better model without HR.

To check this intervention subjectively, we added the trust difference between before and after interaction as a covariate in ANCOVA to check its potential influence on perceived persuasiveness. The results indicated that there is no statistically significant difference between adjusted means of persuasiveness with regard to trust difference (p=.654, effect size .004). In sum, although the trust in robot was increased after the interaction, this factor did not significantly influence the persuasiveness.

We further checked if personal preferences (i.e. liking trivia games or cinema) has affected the results. From the objective perspective, the results indicated no significant effect of liking cinema (Wald(4)=4.176, p=.383). However, liking trivia game turned out to be a good predictor of behavior (Wald(4)=9.671,p=.046). In other words, the more the participants liked trivia games lead to a higher likelihood of accepting the offers (Wald(1)=5.594,p=.018). From the subjective perspective, we performed ANCOVA with liking quiz as a covariate. The results indicated that this covariate does not significantly predict the dependent variable, i.e. persuasiveness of the robot (p=.827, effect size: .000)

Furthermore, earlier research indicated that when there is a power match between the persuader and the persuadee, higher persuasion is achieved [120]. To investigate this effect, we labeled the participants as high/low power based on their PSP (Personal Sense of power) scores (the ones scored higher than the average were labeled as high power). Also, we labeled the robot as being high/low power based on the scores they associated with the reward social power (higher than the average score was labeled as high). Then we checked if a power match exists between the participant and the robot and included it as a covariate in an ANCOVA analysis[2]. The results indicated that the covariate does not adjust the association between the predictor and outcome variable.

Finally, in [57] Ghazali et. al. considered the total number of accepted offers as an indicator of compliance. Similarly, to check if this feature is a predictor of behavior, we applied an ANOVA test and the results indicated a significant difference among the four conditions (F(3, 114)= 4.682, p=.004, $\eta^2 = .110$). Post hoc comparisons using the Tukey HSD test indicated that the mean score for the LR condition (M=1.633, S.E.=.183) was significantly higher than NR condition (M=.821, S.E.=.189). Similarly, the mean score in HR condition (M=1.600, S.E.=.183) was significantly higher than NR. This results endorse the verification of H5, or the presence of robot has significantly affected decision making of the participants.

From the subjective point of view, the result of a Pearson correlation test indicated a strong significant and positive correlation between the total number of accepted offers and perceived persuasiveness. Particularly, the higher the persuasion perceived the higher the number of accepted offers (r(118)=.679, p=.000). Further analysis indicated that this correlation is stronger among LR (r(30)=.795, p=.000), then 0R (r(30)=.781,p=.000), then HR (r(30)=.490, p=.006) and the least for NR (r(28)=.443,.018). This finding is inline with their finding, that when the persuasion is stronger, the compliance decreases due to potential reactance.

7.3.3 Analyzing the Game Log

Apart from the data obtained from the questionnaire, we analyzed the game logs to further investigate how the participants acted and made decisions during the game. In this subsection, we investigate a number of these features.

One of the game features that might influence decision making of participants to accept/reject the offer, is the remaining scores they require to achieve a cinema ticket. Adding this feature to the GEE model indicated that it is a good predictor of the behavior (Wald(1)=6.386, p=.012) and the test of goodness of fit showed a decrease (470.024 vs. 476.080) in the Quasi Likelihood under Independence Model Criterion (QIC) meaning that it is a good predictor of decisions. Also, the results indicates that an increase in distance to ticket increases the likelihood of compliance (log odds: B: .116$\pm$.0458, Exp(B)=1.123).

Another fruitful feature of game logs might be the collected score in each trial. The test of model effects indicated that this is also a good predictor of decision making (Wald(1)=27.409, p=.000). Also, the goodness of fit test showed a decrease in QIC (451.646 vs. 476.080) meaning that a model using this factor as a predictor is a better fit to the data. On the other hand, further analysis indicated that an increase in the score, leads to lower likelihood of acceptance, i.e. the lower score they collected, the higher probability of acceptance the offer in the next trial (odds: -.444$\pm$.0848; EXP .642).

Another game feature that might be informative is the cumulative score or the overall score they gained before making a decision. The test of model effects indicated it as a good predictor (Wald(1)=9.760, p=.002) and the goodness of fitness test showed a decrease in QIC (460.854 vs. 476.080) further endorsing it is a good predictor. However, the negative log odds, i.e. -.226$\pm$.0724 indicated that people

[2]We would like to highlight that scores of power match were equal in case of median- and mean-split.

Table 7.5: GEE model summary for decision change direction

Conditions	Odds	Wald test result
	trial 1	
LR vs. NR	odds:0.758	Wald(1)=3.983,p=0.046
HR vs. NR	odds:0.875	Wald(1)=5.655,p=0.017
0R vs. NR	odds:0.758	Wald: 4.212,p=0.040
	trial 2	
	No significant differences found	
	trial 3	
LR vs. 0R	1.253	Wald(1)=8.860,p=0.003
HR vs. 0R	1.022	Wald(1)=4.996,p=0.025
NR vs. 0R	1.232	Wald(1)=4.326,p=0.038

having collected high scores are less likely to accept the suggestions.

In this game, the participants could change their mind and select options contrasting their stated preferences during the game. This feature gives them an opportunity to cheat the robot! In other words, as the robot always offered the opposing option (the option that participants did not choose), they could initially select the option that they prefer less, and accept the suggestion of the robot to answer their most preferred option (plus receiving points in persuasion scenarios). With this regard, it is not unlikely to observe less cheating on the first trial, when the participants are not aware of robots' function. But the cheating might reach to its maximum on the third trial when the participants become more familiar with robot's function. Analyzing the game log indicated that only one person selected his non-preferred choice and change to the preferred option in all trials. Also, two people did this both in the second and the third trials, however, one of them belonged to 0R condition (might do this randomly), and one belonged to NR. Hence, although the participants had a chance to cheat the robot, only two persons did this.

Finally, we investigated how the participants made decisions considering their initial preferences. With this regard, they could change their opinion in three possible directions: not accepting the offer or no change, accepting the offer and select a less-desirable choice, accepting the offer and select a more-desirable choice. Having a repeated test, we applied GEE model using change directions as responses and trial and conditions as predictors. Table 7.5 summarizes the findings. In sum, we can infer that, in the first trial, the robot was effective and all conditions with robot were higher persuading, even the one with 0 reward. In other words, on the first trial the robot could persuade the users to change their initial selections significantly more frequently than the NR condition. On the second trial the differences between groups were not significant. There is not enough evidence to make any conclusion on this trial. However, on the third trial, all the conditions are higher than 0R, meaning that the reward has an effect on decision making (manipulation checked). Here is the only place were NR become more persuading than 0R. And when the robot is not giving any reward and does not have any power on the participants, over a repeated interaction it acts less persuading over time in comparison with a computer application giving reward constantly (Figure 7.5). A potential reason might be the differences in their selection during the game and what they stated in the initial questionnaire. As we mentioned earlier, it is not probable that

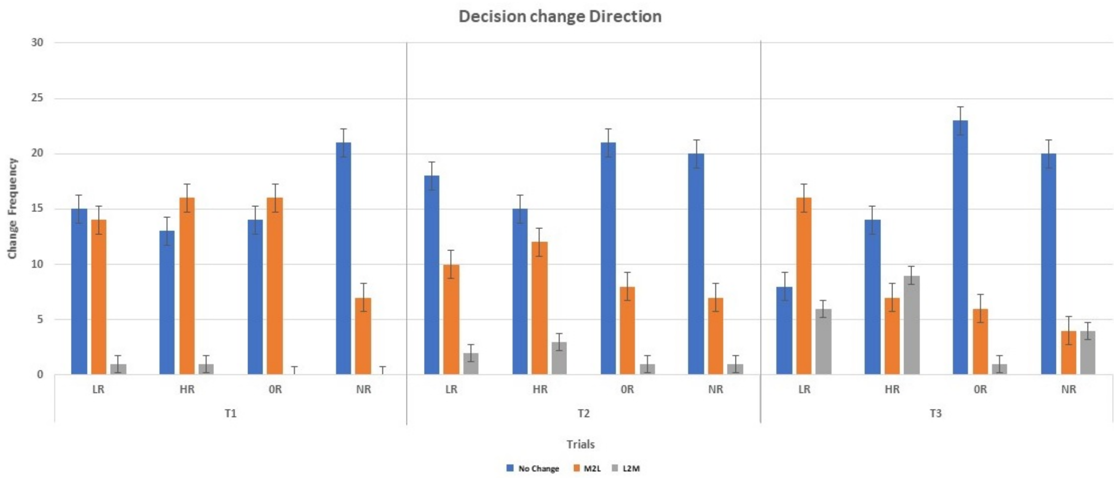


Figure 7.5: Study 3 - Direction of decision change
In this figure, blue bars represent no change (rejecting the offer), M2L stands for changing from a More favorable choice to a Less favorable one, and L2M represents the contrary.

the participants wanted to cheat the robot. They might have not paid enough attention while they were filling out the preference questionnaire, but when it came to reality in the game, they paid more attention to what they refer more.

In the LR condition, people were struggling to get at least one ticket, while in the HR condition, people mostly received one ticket and were struggling to achieve another ticket. Maybe in LR the one extra point could help them to win at least one cinema ticket, but in HR there was a different situation.

7.3.4 Qualitative Analysis

To further investigate and learn more from the data we looked from a closer view by revisiting the data qualitatively. With this new view we aim to interpret and describe data to find and understand potential patterns of behavior in decision making of the participants.

Qualitative analysis is not a new concept in HRI and earlier studies have investigated the finding using qualitative techniques with different goals. For instance, in [108] the authors have used the directed approach [87] for coding the interviews to measure how much the users could recall or learn from the interaction with the robot.

Here, we approached the qualitative analysis to better understand their motivation behind their decision makings. In our analysis, we use a combination of the conventional and the summative approach proposed by [87] to investigate how people experienced the interaction with the robot and made decisions under different circumstances and conditions of the study. The open-ended question aimed at exploring participants' unique perspectives and motivation behind their decision making.

As the result of a qualitative analysis is directly dependent to the coding scheme used, three researchers contributed in coding (data retention) to have a more credible and reliable coding scheme [143]. Initially, the three researchers labeled the data freely and individually to achieve categories as general as possible and not biased by each other. We used the WebQDA software [3] to code and label the open-

[3] https://www.webqda.net/

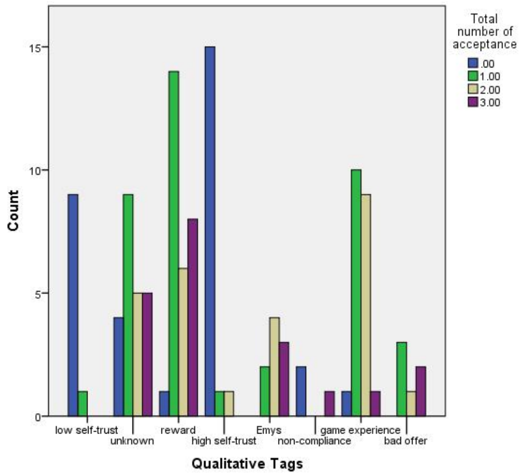


Figure 7.6: Study 3 - Distribution of total number of accepted offers over Qualitative Tags

ended question. Then we discussed and compared the labels to enhance the validity of our coding.

Each of the three researchers identified 13, 11, and 9 categories (listed in Table 7.6) in the first phase. To determine the trustworthiness of codes we further discussed them through checking consistency and reasons for inconsistency. We reached an agreement of 8 categories in the end (low self trust, Unknown, reward, high self-trust, Emys, non-compliance, good offer, Game Experience). In the next step we re-labeled the data using this coding scheme with the goal of minimizing inconsistency. The inter-rater reliability between each two researchers gained K=.653 (p=.000), K=.515, p=.000, and K=.476, p=.000. After another discussion we agreed to use one of the ratings with the highest inter rater reliability.

1. There is a statistically significant difference in the number of accepted offers between people categorized with regard to different tags (f(7,177)=10.160, p=.000). People who had high self-trust were less compliant. These people who valued their own knowledge than the robot's offer tended to reject more than the rest (Figure 7.6).

2. Participants' perception of the robot was significantly affected by their statement regarding robot persuasiveness (F(7,117)=7.924, p=.000) but no difference exists in RoSAS scores (Warmth: F(7,117)=.787, p=.599, Competence: F(7,117)=.678, p=.690, Discomfort: F(7,117)=.559, p=.787).

3. The distribution of the tags is significantly different in different conditions (X^2(21)=41.248, p=.005), as depicted in the Figure 7.7. A Bonferroni post-hoc tests indicated that these differences stand out regarding the game experience tag among HR vs. NR and HR vs. 0R. Specifically, people in HR condition indicated that they accepted the offer to have a higher game experience by accepting the challenge to answer their less-desirable choice. And this higher number (11 in HR and 1 in NR and 1 in 0R) is statistically significant.

Apart from this quantified reasoning, Figure 7.7 depicts that LR and NR (conditions the same reward level, i.e., 1 extra point) have the same number of low self trust, and all conditions have the same number of reward, except 0R. Furthermore, as discussed earlier, the quantitative results could not verify H1 (i.e., no significant difference between high reward and low reward regarding decision making). Interestingly,

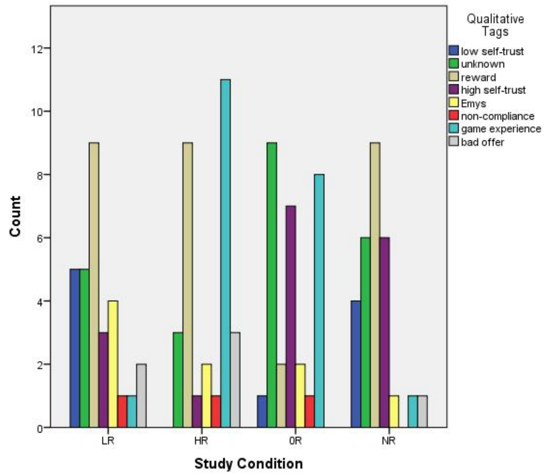


Figure 7.7: Study 3 - Distribution of Qualitative Tags over conditions

we observe a higher number of people with low self-trust in the first condition (LR). The responses to the open-ended question indicates that people in the high reward condition (HR) were less compliant and that could be related to the fact they wanted a challenge, and maybe they were more risk-prone because they did not have to worry too much about the ticket (they had more freedom to change or not and still secure the ticket).

Also, a closer look to the data indicates that in the first control condition (0R) on the first trial, people tend to accept, because they expected to receive some points. For instance, an individual indicated that "I hoped it asks easier question when I changed the subject." But when they received none, they stopped accepting the offer (for example, as indicated by the individual "When I did, I did not change anymore"). So, after not receiving any points, they stopped accepting the offer.

Overall, the qualitative analysis opened up new insights toward the data. As "no qualitative methodology is exclusive" [143], we do not claim that the coding we used is the 'right" and the only applicable scheme. Although three researchers coded the sentences individually and in group to enhance content validation; However, as one's coding changes over time [143] there might be other interpretation of qualitative results. Hence, these findings should be considered cautiously. Furthermore, apart from the coding process, some participants might be shy to directly indicated they wanted the ticket or the reward.

7.4 Discussion

In this section, we presented the result of a user-study performed to investigate the effect of different levels of social power (particularly reward social power) and repeated interactions on persuasion. We hypothesized that higher level of social power would lead to higher persuasion, also having a fixed level of social power this effect would not decay over time. The result of this study did not verify the former neither subjectively nor objectively, and hence, we could not conclude if the increase in power leads to higher persuasion. This finding is somehow similar to the results of the second study, in which the

Table 7.6: Study 3 - Primary coding scheme by 3 researchers

Researcher 1	Researcher 2	Researcher 3
challenge	Reward	Bad Offer
curiosity	Self trust	Challenge
Emys	Low self trust	Confused
fun	Challenge	Curiosity
lack of knowledge	Game experience	Good Offer
low reward	Good offer	More Credit
misunderstood	non compliance	Persuaded
no reason	Emys	Random Choice
distrust	Unknown	Self-Trust
points	Other	
risk	Low reward	
self-trust		
unknown		

increase in rating (that indirectly increased the level of reward) did not lead to any significant difference in decision making of participants. Hence, we may conclude that persuasion does not have a linear relationship with the level of power exerted. This is inline with recent research that indicated a nonlinear relationship between power and persuasion [42].

On the other hand, Ghazali et. al. endorsed that exerting a strong persuasion attempt acts negatively and hence causes reactions and leads to low compliance [60]. Also, they indicated reaction is associated with higher negative cognition and feeling of anger, which might be equivalent to a higher score of discomfort dimension of the RoSAS questionnaire. However, our results did not lead any significant differences in the score of discomfort for people who rejected more frequently comparing to the others (ANOVA: $F(3,114)=1.330$, $p=.268$). In this case, although in HR condition the persuasion was stronger, reactance has not happened. In other words, the rejection was not due to reaction felt.

Hence, our study verifies that power and persuasion do not have a linear relationship, however, further investigation is required to determine this nonlinear relationship. Also further evidence is required to assess the reactance threshold. Apart from this, another potential reason for this insignificant finding might be the small difference between the scores in LR and HR conditions. Although we considered the higher reward to be more than half of the maximum potential achievable scores (3 out of 5), participants might have valued this extra score different from our expectations. A clear information about the state of their mind might be a clue to interpret the results.

Further, the results lead to contradicting findings regarding the latter, i.e., repeated interactions. Specifically, although we expected that the effect of power on persuasion remains unchanged over a repeated interaction, this hypothesis was verified only in two conditions, particularly in HR (high reward) and NR (no robot). In case of 0R or the first control condition, people tend to accept the offer less frequently at the third trial. When they were not gaining any scores for changing, they trusted their own knowledge and did not accept. Hence, not using any sort of power strategy, the robot did not have any persuasive power and people did not comply with the request. However, unlike our expectation, in LR condition, using the same level of power, the robot gained higher persuasion at the end. Interestingly, this finding is inline with Ghazali et al. that the robot with mid-level of persuasion power was more successful than high-power or not robot. However, our result indicated no reaction in HR. We argue that this inconsistency may be due to the value of the reward that the participants associated with in each trial. In other words, it seems possible that the value of the reward was not equal in all conditions. That is to say, when the participants were more near to gaining a cinema ticket, a single score might have a value more than one score in the first trial when they are far away from getting a ticket. As an example, imagine a participant needing only one score to gain a cinema ticket. This one single score means more to him/her, comparing to a person needing 5 scores.

Further, unlike what we hypothesized, people new to robots did not show significantly different results compared to the others. This finding is contradicting to Study 2 findings. A possible explanation for this might be the small sample size of the previous study. Around one fourth of the sample of study two were new to robots and they mostly fell in the same condition. Similar to the previous study, 33% of sample had already interacted with the robots. However, not only the sample size was doubled, but also the sample was more uniformly spread in the groups (each group had around 70 percent people new to robots, except in the RL group that 57 percent of the sample were new to robots).

Unlike the two previous experiments, we did not find any significant differences in the perception of the robots. In the first study, the two robots used completely different dialogues in their interactions with the participants. Also, in the second study, the robot used two different strategies and hence the dialogues used in each condition were different in the persuasive strategy. However, in this study the difference between the conditions was minor and only one single strategy was used in the persuasion conditions. Also, the reward does not increase likeability of the robot.

Finally, we considered the fifth hypothesis (H5) to investigate if the presence of the robot has any effect on the persuasion. Specifically, one might argue that since the robot has no physical interaction with the participants, the persuasion is gained only due to the scores that people receive. In other words, a sole application would do the same job. This is the main reason for adding the fourth condition (NR). However, the results indicated that this argument is not true and the robot's presence leads to another channel of persuasion due to its social presence. Hence, the sense of presence of the robot should not be neglected in this case. This finding is inline with Ghazali et. al. [60]. Although their results indicated that in one condition, i.e., low psychological involvement, the increase in social agency did not influence compliance, but in another condition, or high psychological involvement, compliance remained the same for medium social agency but drooped for high social agency condition. But our finding shows a different trend, when there is no robot, we have achieved significantly less compliance. A potential difference might be that in our study the robot was present, but turned off. Although most of the participants thought "Emys" was the application, we made the expectation that they will interact with the robot later on in another phase shortly after finishing the current tasks. Importantly, we would like to highlight that the results indicated no significant difference among the two control groups, i.e. 0R and NR, in which there was no manipulation.

7.5 Summary of Findings

The findings of this study are four-fold: first, inline with other studies the results indicated that an increase in reward social power (and more generally a stronger persuasion) acts negatively and does not necessary lead to a higher compliance. Second, a prior interaction with robots does not influence the decision making of the participants, unlike what we observed in the second study in which might have been due to the smaller sample size and using a single persuasion attempt. Thirdly, over a repeated interaction the compliance might change due to the specific circumstances (either the study condition, or the user's valuation of the reward) of the study. However, further evidence is required to determine how these circumstances affects decision makings. And finally, the qualitative analysis of contextual data gathered in the study revealed new insights to the data. For instance, people with high self-trust were less compliant to the robot.

7.6 Limitation and suggestion for future studies

One limitation of this study is the use of the questionnaire only before and after the study. In other words, we do not have enough information about the user at each single trial. Hence, we could not measure the subjective measures (robot perception regarding persuasiveness or RoSAS). Further, we did not have enough information how they perceived the trustworthiness of the robot on the second trial.

Another limitation of this study might be the cumulative scores over the three trials. As discussed earlier, 1 single score might have a different meaning depending on the the scores collected so far. So, maybe at each trial we can reset the scores, so that the next round would be independent from the distance to score. Or, we may have a large pool and compare people with the same amount of remaining scores separately.

Likewise any other self-report measure, the primary questionnaire asking about the preference might not be a good measure of users' preferences. In fact, some people selected their less favorable choice initially and indicated in the open-ended question that they did not answer the question carefully. Hence, considering that there is a cheating incentive, we cannot make sure if they really selected their preferences carelessly or they decided to cheat.

Although discomfort is supposed as an indicator of negative cognition, but it might not be a good predictor of reactance. For instance, different evidences indicated less compliance in HR than LR. This might have happened due to reactance to robot's suggestion using a higher level of power or that might be due to the remaining score to get a cinema ticket. As we discussed earlier, the analysis of the game log indicated that at the third trial, people in LR accepted the robot's suggestion more frequently than HR. The study is limited by the lack of information on reactance and a better measure is required.

A future study could assess the effect of different power levels to indicate the level threshold in which reactance happens. In other words, considerably more work will need to be done to determine the relationship between power level and persuasion with regard to reactance.

Further research could also be conducted to determine the effectiveness of behavioral analysis of the participants using the recorded videos. Apart from the contextual data that we analyzed earlier, these behavioral cues could enrich the qualitative analysis.

In final words, these findings provide the insights for future research that reward social power endows persuasiveness to robots. Further work needs to be done to establish whether other power bases are effective in persuasion.

Chapter 8

Summary and Conclusion

In this chapter, we first highlight the key aspects of this document and summarize our contributions. Then we propose the potential future extensions of this dissertation.

8.1 Contribution

Social Power has long been a question of great interest in a wide range of fields including social psychology, politics, organization, etc. Recent advancement in Computer Science and recent social applications calls for attention to social power dynamics. With this aim, we integrate social power in the interaction of social agents with human users. To do so, we approached the problem first, by conceptualizing social power to equip intelligent agents with the ability to process and reason about the social power of other agents. We believe that this capability enhances the social intelligence of such agents and leads to more believable scenarios.

With this aim, we focused on theoretical issues for agent modeling aiming at increasing intelligence and therefore believability of agents that engage in social interaction. In doing so, we extended a previously proposed model of social power inspired by SAPIENT, based on a well-known theory of social power proposed by French and Raven. The main difference between the current model and the previous version is as follows:

1. The model we propose is purely based on French and Raven [54] theory of social power. Hence, personal characteristics and motivational forces have been removed to decrease the complexity of the model. Although one may suspect that neglecting these variables may threaten the generality or expressivity of the model; however, we argue that the model presented here is based on a validated, verified and highly referred model which is designed by experts in this field. Hence, the generality of the presented model is justified regarding the base model's validity. However, running a user study is required to further to make sure such simplifications do not deteriorate the performance comparing to SAPIENT.
2. An important factor that seems missing in SAPIENT is the "induction" or the ability of the actor to cause valences in the target to influence him/her more.

3. In case of legitimate power, SAPIENT only accepts a limited number of norm types, including position, reciprocity, action commitment, and equity. While, based on French and Raven theory, the value force stems from this base of power is only dependent on the internal values and the induction exerted from the actor. Hence, here, not only we do not limit the potential norms to a specific set, but also we simplify the model by focusing on the power resulted from the legitimacy of the roles of the two agents.

4. In case of referent power, in SAPIENT two types of referent power have been identified. Here, we have simplified modeling the value force of this type of power by merging the two types, leading to a simpler model.

Considering these improvements, we expect the model to work with less complexity and be more applicable facing different situations. Since, first, the model considers other factors introduced in French and Raven theory which are missing the SAPIENT. Second, in the case of legitimate power, SAPIENT limits the number of norm types. But in our model, we keep it in a general so that other models of norms can be applied in a prior step. We keep the focus of our model only on the power bases, not peripheral events such as specific norms triggered by the situations.

These simplifications ease the authoring process of the model. To be more specific, in the current version, due to the simpler definition of the model, the initialization of variables has been made easier. As an example, considering the legitimate power, authoring SAPIENT requires defining personality traits, motivations and personal tendencies such as achievement, power and affiliation. If these values were not initialized, the system would attribute random numbers to them. However, in the current version, to investigate potential differences in legitimate power, we just need to initialize variables which are direct influence value forces with regard to this type of power, such as the group shared between the two agents, their role and their powers resulted from their roles.

Overall, by this conceptualization, we identified the important bases of social power and the factors influencing their strength. Having these factors in mind, the major achievements of the present work is the operationalization of social power in designing persuasive agents. Using social power bases, we attempt to design social robots, a specific case of social agents, equipped with social power bases. We selected social robots due to their physicality and the higher sense of presence comparing to virtual agents. We operationalized these power bases within persuasion tasks and attempt to investigate this potential application of social power in human-agent interaction, i.e. persuasion. The link between power and persuasion, as well as the recent application of persuasive technology, motivated us to investigate this link further. We designed three persuasive strategies inspired by social power, particularly expert, reward, and coercion.

Together the results of these studies provide important insights into persuasion in HRI. We argue that our contribution advances the study of robot persuasion by testing new factors (the social power strategies) that may affect persuasion effectiveness. In this direction, our contribution are as follows:

- We identified that the use of social power is effective for persuading people using social robots.

- We investigated this effect using an incentivized real choice and non-imaginary tasks that increases the external validity of the design.
- We used different within- and between-subject studies as well as mixed-designs and investigated the power-persuasion link both objectively and subjectively in the three studies.
- We concluded that one strategy could influence the users both objectively and subjectively. And these two channel of persuasion might not happen both at the same time (as observed in the first study).
- We argue that social rewards can be effective at persuading users and, unlike material rewards, they are unlimited and always available at a lower cost.
- We found that to achieve significantly different perception of the robot in case of warmth, competence and discomfort, the robot dialogue and social cues needs to be different. Minor differences in few dialogue sentences might not lead to a high difference in these scores (as observed in the second study).
- People who are new to robot might be affected by the novelty effect and this threatens the external validity of results. In this case a longer interaction might war off this effect (Study 2).
- Having a fixed level of social power, the effect of power on persuasion does not decay over repeated interactions (Study 3). The effect might become stronger depending under specific circumstances.
- An increase in the level of power does not linearly give rise to persuasion (Study 3).
- The social presence of a robot increases the chance of gaining higher persuasion (Study 3).
- We showed that the use of social power strategies (expertise, coercion and reward) increases robots' power to influence persuasion outcome.
- We considered both the role of the persuasion actor (social robots) and the persuasion target (human participants) in the success of the persuasion. Hence, our approach has the promise of capturing the dynamic effects of actor and target characteristics on persuasion outcomes.
- Qualitative analysis of the data and using open-ended questions opens up further insight to the that might not be easily interpreted using questions with predefined answers.
- Our findings have implication for robotic persuasion in a consumer choice setting.

In final words, our contributions provide new empirical findings and design implications for using humanoid social robots in compliance gaining and behavior change context. Specifically, the link between power and persuasion investigated in this thesis may contribute in addressing some HCI/HRI research problems. The first and foremost implication would be in the decision making processes of intelligent agents (either a virtual character or a robot); for instance, when an agent faces a user with higher power, it would be more believable if it complies with the request of the user (not necessarily always). Furthermore, adapting the interaction by considering the level of power would lead to more believability. For instance, the agent with a low power state should utilize messages with more warmth than competence. On the contrary, high-power communicator agents needs to generate messages with more competence information to gain higher persuasion.

In other words, the agents may use power as a heuristic for persuasion to induce greater compliance in the target. In addition, based on what was alluded to, the sense of power affects action and orientation which directly affects believability of agents. Specifically, an increase in power fosters sense of responsibility, and to have a more believable agent this increase needs to be evident in the behavior and actions of the agent. Furthermore, it has been shown that level of power of a target affects his/her attitude after an argument. This difference should be addressed to gain higher believability.

8.2 Limitation and Future Work

In this section we discuss some of the general limitations in our research studies. Throughout the gradual progress of the thesis, we aimed to overcome the limitation of different studies we faced in the succeeding study. For instance, the first study design was limited by the absence of a sophisticated control condition. Or the small sample size of the second study did not allow achieving significant findings. Further, an additional uncontrolled factor in the first and second study is the unbalanced gender of the participants in different conditions. We controlled this in the third study by balancing an equal number of participants in each conditions. However, since the sample was a random participation of volunteers we could not reach 50-50 distribution.

Also, one source of weakness in the third study which could have affected the measurements of the subjective finding was the absence of measurement after each trial. Although we attempted to overcome this by analysing qualitatively the contextual responses of the participants, however, not all the participants described their entire experience over all the three trials.

Another limitation of the studies might be a full understanding of participants' thought and perceptions. We added an open-ended question to the second study to further investigate this, but as answering the question was not mandatory a number of subjects skipped it. We made the open-ended question mandatory in the third study, still some participants skipped it by putting irrelevant or too short answers.

Finally, likewise any other laboratory experiments, our results might lack in ecological validity suffer from generalization to other situation due to the experimental effects. Furthermore, the sample size of the participants in the studies was relatively small and hence the results may not apply to a broader populations.

8.2.1 Future Work

Our findings provide the following insights for future research. As discussed in Chapter 4, from the theoretical point of view, work needs to be done to further improve the work, including

- By the simplification, we referred to the number of parameters and implementation complexities. However, further examination of the complexity of the work, i.e. comparisons of time complexity and space complexity between the two models is not applicable at this moment.
- It is necessary to perform user-studies to determine if agents with this model were more socially believable than those without it.
- Works need to be done to examine if removing motivational factor decrease expressiveness of the model comparing to SAPIENT.
- Another interesting avenue of research might be modeling the dynamics and evolution of social power over time.
- The model we proposed only considers two agents: actor and target. It might be of interest to investigate how easy would it be to scale this model up to more than two agents.
- As power dynamics are bidirectional, further studies need to be carried out in order to put the power in the individual and investigate this type of social interaction from the other perspective. This would also be a fruitful area for further work.

Apart from the theoretical viewpoint, our user-studies have thrown up many questions in need of further investigation. First and foremost is to add more qualitative approaches to better understand attitudes and behaviors of the subject. We suggest running systematic interviews after the study using direct or indirect questions. This is an intriguing issue which could be usefully explored in further research.

Further research might explore the effect of social power on persuasion in group and social collectives. A considerable amount of literature exist on grouping people and robots. However, less is known about the dynamics of social power within groups of humans and robots.

On the other hand, as discussed earlier, power exists in bidirectional relationships. In a next step, further research might explore the problem by putting the power within the user. This would ease designing scenarios that are more feasible with less ethical issues. For instance, having a robot with legitimate power might not be as believable as a legitimate human user. This would be a fruitful area for further work to design social robots capable of processing social interactions that deal with social power in the next step.

Furthermore, it is necessary to investigate the ethical quandary of persuasive robots. Digital technology has changed the nature of persuasion in several key respects. It has increased complexity, blurring lines among information, entertainment, and influence. With the advent of technology, new tools for persuasion has been provided, for instance social agents. This gives rise to the ethical concerns about the use of persuasion that needs to be considered within the new persuasive technology. Persuasive and powerful robots could support and foster the human user's interests (e.g., in therapy sessions, diet monitoring, or suicide prevention) but could also deceive and manipulate the user (e.g., in sales and political propaganda). Persuasive technologies have demonstrated their effectiveness to negatively impact a user's behavior and generate addictions towards current social technologies. Additionally, a recent study investigated the security risks of persuasive social robots that aim to manipulate people [167]. Using three proof of concepts, the results suggested that the over-trust in robots could provide a risk of being misused and to hack into sensitive information. This does not lie in the scope of this work, however, a future study worth investigating it due to its importance on human societies.

And last but not least, robotic persuaders leading the persuadee might be considered as a specific case of recommender systems. For instance, considering that the "Expert" robot is providing an explanation of why the human should follow its persuasive advice, it would be interesting to put the work presented in this thesis in context of explainable AI; i.e. explainable recommender systems. This would be another important practical implication for future practice.

Bibliography

[1] A. Agnihotri and H. Knight. Persuasive chairbots: A robot recruited experiment. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 700–702. IEEE, 2019.

[2] S. Agrawal and M.-A. Williams. Robot authority and human obedience: A study of human behaviour using a robot security guard. In *Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 57–58. ACM, 2017.

[3] I. M. Albaina, T. Visser, C. A. Van Der Mast, and M. H. Vastenburg. Flowie: A persuasive virtual coach to motivate elderly individuals to walk. In *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–7. IEEE, 2009.

[4] G. T. Allison. *Essence of decision: Explaining the Cuban missile crisis*. Little, Brown and Company, 1971.

[5] E. Anagnostopoulou, B. Magoutas, E. Bothos, J. Schrammel, R. Orji, and G. Mentzas. Exploring the links between persuasion, personality and mobility types in personalized mobility applications. In *International Conference on Persuasive Technology*, pages 107–118. Springer, 2017.

[6] C. Anderson and J. L. Berdahl. The experience of power: Examining the effects of power on approach and inhibition tendencies. *Journal of personality and social psychology*, 83(6):1362, 2002.

[7] C. Anderson, O. P. John, and D. Keltner. The personal sense of power. *Journal of personality*, 80(2):313–344, 2012.

[8] S. Andrist, E. Spannan, and B. Mutlu. Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 341–348. IEEE, 2013.

[9] Y. Bachrach, E. Markakis, E. Resnick, A. D. Procaccia, J. S. Rosenschein, and A. Saberi. Approximating power indices: theoretical and empirical analysis. *Autonomous Agents and Multi-Agent Systems*, 20(2):105–122, 2010.

[10] L. Bechade, G. D. Duplessis, and L. Devillers. Empirical study of humor support in social human-robot interaction. In *International Conference on Distributed, Ambient, and Pervasive Interactions*, pages 305–316. Springer, 2016.

[11] A. Bogdanovych, T. Trescak, and S. Simoff. What makes virtual agents believable? *Connection Science*, 28(1):83–108, 2016.

[12] G. Bohner and N. Dickel. Attitudes and attitude change. *Annual review of psychology*, 62:391–417, 2011.

[13] G. Bohner, H.-P. Erb, and F. Siebler. Information processing approaches to persuasion. integrating assumptions from the dual-and single-processing perspectives. *Attitudes and attitude change*, pages 161–188, 2008.

[14] M. Bratman. *Intention, plans, and practical reason*. Cambridge University Press, 1987.

[15] M. Brauer and R. Y. Bourhis. Social power. *European Journal of Social Psychology*, 36(4):601–616, 2006.

[16] C. Breazeal. Socially intelligent robots. *interactions*, 12(2):19–22, 2005.

[17] C. Breazeal, A. Takanishi, and T. Kobayashi. Social robots that interact with people. *Springer handbook of robotics*, pages 1349–1369, 2008.

[18] P. Briñol, R. E. Petty, G. R. Durso, and D. D. Rucker. Power and persuasion: Processes by which perceived power can influence evaluative judgments. *Review of General Psychology*, 21(3):223, 2017.

[19] B. Bruno, N. Y. Chong, H. Kamide, S. Kanoria, J. Lee, Y. Lim, A. K. Pandey, C. Papadopoulos, I. Papadopoulos, F. Pecora, et al. Paving the way for culturally competent robots: A position paper. In *2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 553–560. IEEE, 2017.

[20] J. T. Cacioppo and R. E. Petty. The need for cognition. *Journal of personality and social psychology*, 42(1):116, 1982.

[21] C. Carabelea, O. Boissier, and C. Castelfranchi. Using social power to enable agents to reason about being part of a group. In *International Workshop on Engineering Societies in the Agents World*, pages 166–177. Springer, 2004.

[22] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner. The robotic social attributes scale (rosas): Development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, pages 254–262. ACM, 2017.

[23] N. Caruana, D. Spirou, and J. Brock. Human agency beliefs influence behaviour during virtual social interactions. *PeerJ*, 5:e3819, 2017.

[24] C. Castelfranchi. Social power: A point missed in multi-agent, dai and hci. *Decentralized AI*, pages 49–62, 1990.

[25] Á. Castro-González, J. Alcocer-Luna, M. Malfaz, F. Alonso-Martín, and M. A. Salichs. Evaluation of artificial mouths in social robots. *IEEE Transactions on Human-Machine Systems*, 48(99):1–11, 2018.

[26] M. Cavazza, C. Smith, D. Charlton, N. Crook, J. Boye, S. Pulman, K. Moilanen, D. Pizzi, R. S. De La Camara, and M. Turunen. Persuasive dialogue based on a narrative theory: an eca implementation. In *International Conference on Persuasive Technology*, pages 250–261. Springer, 2010.

[27] S. Chaiken. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology*, 39(5):752, 1980.

[28] X.-P. Chen, X. Yao, and S. Kotha. Entrepreneur passion and preparedness in business plan presentations: a persuasion analysis of venture capitalists' funding decisions. *Academy of Management journal*, 52(1):199–214, 2009.

[29] V. Chidambaram, Y.-H. Chiang, and B. Mutlu. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 293–300. ACM, 2012.

[30] M. Chromik, B. Carlmeyer, and B. Wrede. Ready for the next step? investigating the effect of incremental information presentation in an object fetching task. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 95–96, 2017.

[31] R. B. Cialdini. The science of persuasion. *Scientific American Mind*, 14(1):70–77, 2004.

[32] C. Clark. Emotions and micropolitics in everyday life: Some patterns and paradoxes of "place.". *Research agendas in the sociology of emotions*, pages 305–333, 1990.

[33] A. Cruz-Maya and A. Tapus. Negotiating with a robot: analysis of regulatory focus behavior. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–9. IEEE, 2018.

[34] S. CUTLIP. Power and persuasion-carter, jf, 1960.

[35] R. A. Dahl. The concept of power. *Systems Research and Behavioral Science*, 2(3):201–215, 1957.

[36] K. Dautenhahn. *Socially intelligent agents: Creating relationships with computers and robots*, volume 3. Springer Science & Business Media, 2002.

[37] K. Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1480):679–704, 2007.

[38] E. Demurie, H. Roeyers, D. Baeyens, and E. Sonuga-Barke. The effects of monetary and social rewards on task performance in children and adolescents: liking is not enough. *International journal of methods in psychiatric research*, 21(4):301–310, 2012.

[39] J. Dias, S. Mascarenhas, and A. Paiva. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Emotion modeling*, pages 44–56. Springer, 2014.

[40] W. Doise and E. T. Mapstone. *Levels of explanation in social psychology.* Cambridge University Press, 1986.

[41] K. Dowding. *Encyclopedia of power.* Sage Publications, 2011.

[42] D. Dubois, D. D. Rucker, and A. D. Galinsky. Dynamics of communicator and audience power: The persuasiveness of competence versus warmth. *Journal of Consumer Research*, 43(1):68–85, 2016.

[43] N. E. Dunbar and J. K. Burgoon. Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships*, 22(2):207–233, 2005.

[44] S. Elias. Fifty years of influence in the workplace: The evolution of the french and raven power taxonomy. *Journal of Management History*, 14(3):267–283, 2008.

[45] C. Elliott and J. Brzezinski. Autonomous agents as synthetic characters. *AI magazine*, 19(2):13, 1998.

[46] R. M. Emerson. Power-dependence relations. *American sociological review*, pages 31–41, 1962.

[47] A. P. Fiske. The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological review*, 99(4):689, 1992.

[48] S. T. Fiske. Controlling other people: The impact of power on stereotyping. *American psychologist*, 48(6):621, 1993.

[49] S. T. Fiske and J. Berdahl. Social power. *Social psychology: Handbook of basic principles*, 2:678–692, 2007.

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

[51] M. Foucault. *The history of sexuality: An introduction*. Vintage, 1990.

[52] L. J. Francis, L. B. Brown, and R. Philipchalk. The development of an abbreviated form of the revised eysenck personality questionnaire (epqr-a): Its use among students in england, canada, the usa and australia. *Personality and individual differences*, 13(4):443–449, 1992.

[53] S. M. Freeman, N. Rebout, and K. L. Bales. Effect of reward type on object discrimination learning in socially monogamous coppery titi monkeys (callicebus cupreus). *American journal of primatology*, page e22868, 2018.

[54] J. R. French, B. Raven, and D. Cartwright. The bases of social power. *Classics of organization theory*, 7:311–320, 1959.

[55] S. Gama, G. Barata, D. Gonçalves, R. Prada, and A. Paiva. A model for social regulation of user-agent relationships. In *International Conference on Intelligent Virtual Agents*, pages 319–326. Springer, 2012.

[56] R. H. Gass and J. S. Seiter. *Persuasion: Social Inflence and Compliance Gaining*. Routledge, 2015.

[57] A. S. Ghazali, J. Ham, E. Barakova, and P. Markopoulos. The influence of social cues in persuasive social robots on psychological reactance and compliance. *Computers in Human Behavior*, 87:58–65, 2018.

[58] A. S. Ghazali, J. Ham, E. Barakova, and P. Markopoulos. Assessing the effect of persuasive robots interactive social cues on users' psychological reactance, liking, trusting beliefs and compliance. *Advanced Robotics*, 33(7-8):325–337, 2019.

[59] A. S. Ghazali, J. Ham, E. I. Barakova, and P. Markopoulos. Pardon the rude robot: Social cues diminish reactance to high controlling language. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 411–417. IEEE, 2017.

[60] A. S. Ghazali, J. Ham, E. I. Barakova, and P. Markopoulos. Effects of robot facial characteristics and gender in persuasive human-robot interaction. *Frontiers in Robotics and AI*, 5:73, 2018.

[61] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, et al. Designing robots for long-term social interaction. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1338–1343. IEEE, 2005.

[62] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.

[63] J. A. Gray. *The neuropsychology of anxiety*. Clarendon Press, 1982.

[64] E. T. Hall and T. Hall. *The silent language*, volume 948. Anchor books, 1959.

[65] J. A. Hall and A. G. Halberstadt. "subordination" and sensitivity to nonverbal cues: A study of married working women. *Sex Roles*, 31(3-4):149–165, 1994.

[66] J. Ham, R. H. Cuijpers, and J.-J. Cabibihan. Combining robotic persuasive strategies: the persuasive power of a storytelling robot that uses gazing and gestures. *International Journal of Social Robotics*, 7(4):479–487, 2015.

[67] J. Ham and C. J. Midden. A persuasive robot to stimulate energy conservation: the influence of positive and negative social feedback and task similarity on energy-consumption behavior. *International Journal of Social Robotics*, 6(2):163–171, 2014.

[68] N. Hanna and D. Richards. Towards a ‘smart’collaborative virtual environment and multi-agent approach to designing an intelligent virtual agent. In *Agent Environments for Multi-Agent Systems IV*, pages 170–187. Springer, 2015.

[69] N. Harvey and I. Fischer. Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational behavior and human decision processes*, 70(2):117–133, 1997.

[70] M. Hashemian. Persuasive social robots using social power dynamics. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 2408–2410. International Foundation for Autonomous Agents and Multiagent Systems, 2019.

[71] M. Hashemian, M. Couto, S. Mascarenhas, A. Paiva, P. A. Santos, and R. Prada. Investigating reward/punishment strategies in the persuasiveness of social robots. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 863–868. IEEE, 2020.

[72] M. Hashemian, M. Couto, S. Mascarenhas, A. Paiva, P. A. Santos, and R. Prada. Persuasive social robots using reward/coercion strategies. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 230–232, 2020.

[73] M. Hashemian, M. Couto, S. Mascarenhas, A. Paiva, P. A. Santos, and R. Prada. Building persuasive robots with social power strategies. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021.

[74] M. Hashemian, M. Couto, S. Mascarenhas, A. Paiva, P. A. Santos, and R. Prada. Persuasive social robot using reward power over repeated instances of persuasion. In *International Conference on Persuasive Technology*. Springer, 2021.

[75] M. Hashemian, S. Mascarenhas, M. Couto, A. Paiva, P. A. Santos, and R. Prada. The application of social power in persuasive social robots. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 564–566, 2020.

[76] M. Hashemian, H. Moradi, and M. S. Mirian. “how is his/her mood”: A question that a companion robot may be able to answer. In *International Conference on Social Robotics*, pages 274–284. Springer, 2016.

[77] M. Hashemian, A. Paiva, S. Mascarenhas, P. A. Santos, and R. Prada. The power to persuade: a study of social power in human-robot interaction. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–8. IEEE, 2019.

[78] M. Hashemian, A. Paiva, S. Mascarenhas, P. A. Santos, and R. Prada. Social power in human-robot interaction: Towards more persuasive robots. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 2015–2017, 2019.

[79] M. Hashemian, R. Paradeda, C. Guerra, and A. Paiva. Do you trust me? investigating the formation of trust in social robots. In *EPIA Conference on Artificial Intelligence*, pages 357–369. Springer, 2019.

[80] M. Hashemian, R. Prada, P. A. Santos, J. Dias, and S. Mascarenhas. Inferring emotions from touching patterns. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–7. IEEE, 2019.

[81] M. Hashemian, R. Prada, P. A. Santos, and S. Mascarenhas. Enhancing social believability of virtual agents using social power dynamics. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 147–152, 2018.

[82] D. Hayes and H. Hexmoor. Social power as an exchangeable resource for distributed multi-agent systems. In *Collaborative Technologies and Systems, 2006. CTS 2006. International Symposium on*, pages 278–281. IEEE, 2006.

[83] S. Herse, J. Vitale, D. Ebrahimian, M. Tonkin, S. Ojha, S. Sidra, B. Johnston, S. Phillips, S. L. K. C. Gudi, J. Clark, et al. Bon appetit! robot persuasion for food recommendation. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 125–126, 2018.

[84] T. Holz, M. Dragone, and G. M. O'Hare. Where robots and virtual agents meet. *International Journal of Social Robotics*, 1(1):83–93, 2009.

[85] Y. T.-Y. Hou and M. Jung. Robots in power. In *Proceedings of Longitudinal Human-Robot Teaming Workshop at HRI18*, pages 325–338. ACM, 2018.

[86] C. I. Hovland, A. A. Lumsdaine, and F. D. Sheffield. Experiments on mass communication. *Studies in social psychology in World War II, Vol. 3*, 1949.

[87] H.-F. Hsieh and S. E. Shannon. Three approaches to qualitative content analysis. *Qualitative health research*, 15(9):1277–1288, 2005.

[88] Y. Imai. The relationship between perceived social power and the perception of being influenced. *Japanese Psychological Research*, 31(3):97–107, 1989.

[89] W. JUa. Power in human robot interactions. *What Social Robots Can and Should Do: Proceedings of Robophilosophy 2016/TRANSOR 2016*, 290:13, 2016.

[90] D. Kahneman, J. L. Knetsch, and R. H. Thaler. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5(1):193–206, 1991.

[91] S.-H. Kang, D. M. Krum, P. Khooshabeh, T. Phan, C.-Y. Chang, O. Amir, and R. Lin. Social influence of humor in virtual human counselor's self-disclosure. *Computer Animation and Virtual Worlds*, 28(3-4):e1763, 2017.

[92] S. K. Kang, A. D. Galinsky, L. J. Kray, and A. Shirako. Power affects performance when the pressure is on: Evidence for low-power threat and high-power lift. *Personality and Social Psychology Bulletin*, 41(5):726–735, 2015.

[93] H. H. Kelley and J. W. Thibaut. *Interpersonal relations: A theory of interdependence*. John Wiley & Sons, 1978.

[94] D. Keltner, D. H. Gruenfeld, and C. Anderson. Power, approach, and inhibition. *Psychological review*, 110(2):265, 2003.

[95] D. Keltner, G. A. Van Kleef, S. Chen, and M. W. Kraus. A reciprocal influence model of social power: Emerging principles and lines of inquiry. *Advances in experimental social psychology*, 40:151–192, 2008.

[96] P. Khooshabeh, C. McCall, S. Gandhe, J. Gratch, and J. Blascovich. Does it matter if a computer jokes. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 77–86. ACM, 2011.

[97] C. Kim and A. L. Baylor. A virtual change agent: Motivating pre-service teachers to integrate technology in their future classrooms. *Journal of Educational Technology & Society*, 11(2):309–321, 2008.

[98] Y. Kim and B. Mutlu. How social distance shapes human–robot interaction. *International Journal of Human-Computer Studies*, 72(12):783–795, 2014.

[99] K. W. Kobberholm, K. S. Carstens, L. W. Bøg, M. H. Santos, S. Ramskov, S. A. Mohamed, and L. C. Jensen. The influence of incremental information presentation on the persuasiveness of a robot. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 302–304, 2020.

[100] G. Kohls, J. Peltzer, B. Herpertz-Dahlmann, and K. Konrad. Differential effects of social and non-social reward on response inhibition in children and adolescents. *Developmental science*, 12(4):614–625, 2009.

[101] N. C. Krämer, N. Simons, and S. Kopp. The effects of an embodied conversational agent's nonverbal behavior on user's evaluation and behavioral mimicry. In *International Workshop on Intelligent Virtual Agents*, pages 238–251. Springer, 2007.

[102] N. C. Krämer, A. von der Pütten, and S. Eimler. Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. In *Human-computer interaction: The agency perspective*, pages 215–240. Springer, 2012.

[103] A. W. Kruglanski and E. P. Thompson. Persuasion by a single route: A view from the unimodel. *Psychological Inquiry*, 10(2):83–109, 1999.

[104] J. Lafferty, P. M. Eady, and J. Elmers. The desert survival problem. *Experimental Learning Methods*, 1974.

[105] J. E. Laird, A. Newell, and P. S. Rosenbloom. Soar: An architecture for general intelligence. *Artificial intelligence*, 33(1):1–64, 1987.

[106] J. Lammers, D. Dubois, D. D. Rucker, and A. D. Galinsky. Power gets the job: Priming power improves interview outcomes. *Journal of Experimental Social Psychology*, 49(4):776–779, 2013.

[107] S. A. Lee and Y. J. Liang. Robotic foot-in-the-door: Using sequential-request persuasive strategies in human-robot interaction. *Computers in Human Behavior*, 90:351–356, 2019.

[108] I. Leite, M. McCoy, M. Lohani, D. Ullman, N. Salomons, C. Stokes, S. Rivers, and B. Scassellati. Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 75–82. IEEE, 2015.

[109] N. Lete, A. Beristain, and A. García-Alonso. Survey on virtual coaching for older adults. *Health Informatics Journal*, page 1460458220926894, 2020.

[110] J. Li. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77:23–37, 2015.

[111] J. Li, W. Ju, and C. Nass. Observer perception of dominance and mirroring behavior in human-robot relationships. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 133–140. ACM, 2015.

[112] N. Liberman, Y. Trope, and E. Stephan. Psychological distance. *Social psychology: Handbook of basic principles*, 2:353–383, 2007.

[113] A. Lopez, B. Ccasane, R. Paredes, and F. Cuellar. Effects of using indirect language by a robot to change human attitudes. In *Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 193–194. ACM, 2017.

[114] M. Mann. *The sources of social power: Volume 3, global empires and revolution, 1890-1945*, volume 3. Cambridge University Press, 1986.

[115] S. Mascarenhas, R. Prada, A. Paiva, and G. J. Hofstede. Social importance dynamics: A model for culturally-adaptive agents. In *International Workshop on Intelligent Virtual Agents*, pages 325–338. Springer, 2013.

[116] M. Mattson. A review of: Perloff, rm (2003).“the dynamics of persuasion: Communication and attitudes in the 21st century ” mahwah, nj: Lawrence erlbaum associates, inc. isbn: 0805840885 (softcover) 46.50;isbn:0805840877(hardcover) 99.95, 416 pages, 2005.

[117] D. C. McClelland. *Human motivation*. CUP Archive, 1987.

[118] C. Midden and J. Ham. Using negative and positive social feedback from a robotic agent to save energy. In *Proceedings of the 4th international conference on persuasive technology*, page 12. ACM, 2009.

[119] D. Modic, R. Anderson, and J. Palomäki. We will make you like our research: The development of a susceptibility-to-persuasion scale. *PloS one*, 13(3):e0194119, 2018.

[120] M. Mourali and A. Nagpal. The powerful select, the powerless reject: Power's influence in decision strategies. *Journal of Business Research*, 66(7):874–880, 2013.

[121] C. Nass, J. Steuer, E. Tauber, and H. Reeder. Anthropomorphism, agency, and ethopoeia: computers as social actors. In *INTERACT'93 and CHI'93 conference companion on Human factors in computing systems*, pages 111–112. ACM, 1993.

[122] C. Nass, J. Steuer, and E. R. Tauber. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78. ACM, 1994.

[123] R. Niewiadomski, V. Demeure, and C. Pelachaud. Warmth, competence, believability and virtual agents. In *International Conference on Intelligent Virtual Agents*, pages 272–285. Springer, 2010.

[124] A. Nijholt. Smart bugs and digital banana peels: Accidental humor in smart environments? In *International Conference on Distributed, Ambient, and Pervasive Interactions*, pages 329–340. Springer, 2016.

[125] S. Okumura, M. Kimoto, M. Shiomi, T. Iio, K. Shimohara, and N. Hagita. Do social rewards from robots enhance offline improvements in motor skills? In *International Conference on Social Robotics*, pages 32–41. Springer, 2017.

[126] S. Oreg and N. Sverdlik. Source personality and persuasiveness: Big five predispositions to being persuasive and the role of message involvement. *Journal of personality*, 82(3):250–264, 2014.

[127] R. Paradeda, M. Hashemian, C. Guerra, R. Prada, J. Dias, and A. Paiva. Fides: How emotions and small talks may influence trust in an embodied vs. non-embodied robot. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 1673–1675, 2017.

[128] R. B. Paradeda, M. Hashemian, R. A. Rodrigues, and A. Paiva. How facial expression may influence the trust in a robot? In *2016 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016.

[129] R. B. Paradeda, M. Hashemian, R. A. Rodrigues, and A. Paiva. How facial expressions and small talk may influence trust in a robot. In *International Conference on Social Robotics*, pages 169–178. Springer, 2016.

[130] G. Pereira, R. Prada, and P. A. Santos. Bases of social power for agents. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 1289–1290. International Foundation for Autonomous Agents and Multiagent Systems, 2013.

[131] G. Pereira, R. Prada, and P. A. Santos. Integrating social power into the decision-making of cognitive agents. *Artificial Intelligence*, 241:1–44, 2016.

[132] G. D. G. Pereira, R. Prada, and P. A. Santos. Social theatre: Showcasing social power aware agents. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1951–1952. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

[133] R. M. Perloff. *The dynamics of persuasion: Communication and attitudes in the 21st century*. Routledge, 1993.

[134] R. E. Petty and J. T. Cacioppo. The elaboration likelihood model of persuasion. In *Communication and persuasion*, pages 1–24. Springer, 1986.

[135] A. S. Rao, M. P. Georgeff, et al. Bdi agents: from theory to practice. In *ICMAS*, volume 95, pages 312–319, 1995.

[136] B. H. Raven. A power/interaction model of interpersonal influence: French and raven thirty years later. *Journal of Social Behavior & Personality*, 1992.

[137] B. H. Raven. The bases of power and the power/interaction model of interpersonal influence. *Analyses of Social Issues and Public Policy*, 8(1):1–22, 2008.

[138] K. K. Reardon. *Persuasion in practice.* Sage, 1991.

[139] B. Reeves and C. I. Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press, 1996.

[140] T. Ribeiro, A. Paiva, and D. Dooley. Nutty tracks: symbolic animation pipeline for expressive robotics., 2013.

[141] T. Ribeiro, A. Pereira, E. Di Tullio, P. Alves-Oliveira, and A. Paiva. From thalamus to skene: High-level behaviour planning and managing for mixed-reality characters. In *Proceedings of the IVA 2014 Workshop on Architectures and Standards for IVAs*, 2014.

[142] T. Ribeiro, A. Pereira, E. Di Tullio, and A. Paiva. The sera ecosystem: Socially expressive robotics architecture for autonomous human-robot interaction. In *Enabling Computing Research in Socially Intelligent Human-Robot Interaction: A CommunityDriven Modular Research Platform*, 2016.

[143] L. Richards. *Handling qualitative data: A practical guide*. Sage, 2014.

[144] D. L. Roberts, M. L. Furst, B. Dorn, and C. L. Isbell. Using influence and persuasion to shape player experiences. In *Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games*, pages 23–30. ACM, 2009.

[145] E. Romero, P. Villar, J. A. Gómez-Fraguela, and L. Lopez-Romero. Measuring personality traits with ultra-short scales: A study of the ten item personality inventory (tipi) in a spanish sample. *Personality and Individual Differences*, 53(3):289–293, 2012.

[146] S. Rossi and P. D'Alterio. Gaze behavioral adaptation towards group members for providing effective recommendations. In *International Conference on Social Robotics*, pages 231–241. Springer, 2017.

[147] M. A. Roubroeks, J. R. Ham, and C. J. Midden. The dominant robot: Threatening robots cause psychological reactance, especially when they have incongruent goals. In *International Conference on Persuasive Technology*, pages 174–184. Springer, 2010.

[148] J. Rouchier. Social intelligence for computers. In *Socially Intelligent Agents*, pages 85–92. Springer, 2002.

[149] D. Rousseau and B. Hayes-Roth. A social-psychological model for synthetic actors. In *Proceedings of the second international conference on Autonomous agents*, pages 165–172. ACM, 1998.

[150] B. Russell. *Power: A new social analysis*. Routledge, 2004.

[151] S. J. Russell and P. Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.

[152] S. Saunderson and G. Nejat. It would make me happy if you used my guess: Comparing robot persuasive strategies in social human–robot interaction. *IEEE Robotics and Automation Letters*, 4(2):1707–1714, 2019.

[153] S. Saunderson and G. Nejat. Investigating strategies for robot persuasion in social human-robot interaction. *IEEE Transactions on Cybernetics*, 2020.

[154] M. Schaerer, L. P. Tost, L. Huang, F. Gino, and R. Larrick. Advice giving: A subtle pathway to power. *Personality and Social Psychology Bulletin*, 44(5):746–761, 2018.

[155] D. Shapiro, K. Tanenbaum, J. McCoy, L. LeBron, C. Reynolds, A. Stern, M. Mateas, B. Ferguson, D. Diller, K. Moffitt, et al. Composing social interactions via social games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 573–580. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

[156] M. Siegel, C. Breazeal, and M. I. Norton. Persuasive robotics: The influence of robot gender on human behavior. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 2563–2568. IEEE, 2009.

[157] J. L. Swasy. Measuring the bases of social power. *ACR North American Advances*, 1979.

[158] G. Taylor, R. Bechtel, G. Morgan, and E. Waltz. A framework for modeling social power structures. *Ann Arbor*, 1001:48105, 2006.

[159] C. Torrey. *How robots can help: Communication strategies that improve social outcomes*. PhD thesis, Carnegie Mellon University, 2009.

[160] C. Torrey, A. Powers, M. Marge, S. R. Fussell, and S. Kiesler. Effects of adaptive robot dialogue on information exchange and social relations. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 126–133, 2006.

[161] J. C. Turner. Explaining the nature of power: A three-process theory. *European journal of social psychology*, 35(1):1–22, 2005.

[162] A. Valitutti and T. Veale. Infusing humor in unexpected events. In *International Conference on Distributed, Ambient, and Pervasive Interactions*, pages 370–379. Springer, 2016.

[163] J. Vilk and N. T. Fitter. Comedy by jon the robot. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 80–80, 2020.

[164] D. Wang, T. Liu, and J. Shi. Development of monetary and social reward processes. *Scientific reports*, 7(1):11128, 2017.

[165] E. V. Wilson. Perceived effectiveness of interpersonal persuasion strategies in computer-mediated communication. *Computers in Human Behavior*, 19(5):537–552, 2003.

[166] K. Winkle, S. Lemaignan, P. Caleb-Solly, U. Leonards, A. Turton, and P. Bremner. Effective persuasion strategies for socially assistive robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 277–285. IEEE, 2019.

[167] P. Wolfert, J. Deschuyteneer, D. Oetringer, N. Robinson, and T. Belpaeme. Security risks of social robots used to persuade and manipulate: A proof of concept study. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 523–525, 2020.

[168] E. Xiao and D. Houser. Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences*, 102(20):7398–7401, 2005.

[169] S. You and L. P. Robert Jr. Human-robot similarity and willingness to work with a robotic co-worker. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 251–260, 2018.

[170] G. Yukl. *Leadership in Organizations, 9/e*. Pearson Education India, 1981.

[171] J. Złotowski, K. Yogeeswaran, and C. Bartneck. Can we control it? autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, 100:48–54, 2017.

Appendix A

Study 1: Full Dialogues

In this scenario, Gleen is Expert and Emys is Joker.

Table A.1: Robot Dialogue in Study 1

#	robot	Dialogue
#1	Expert	<Gaze(person)>Dear + namePlayer + , my name is Gleen. Welcome to our coffee testing program!
#2	Joker	<Gaze(person)>Hello + namePlayer +, my name is Emys. Glad to see you here <Gaze(person3)>
#3	Expert	<Gaze(person)>Hey + namePlayer +, do you like coffee? <Gaze(Joker)>
#4a	Joker (Positive)	<Gaze(person)>Great, I also like coffee. That’s why I am working here. Hih hih!
#4b	Joker (Negative)	<Gaze(person)>Oh, you don’t? But I do love coffee. That’s why I’m working here. Hih hih!
#4c	Joker (Neutral)	<Gaze(person)>Well, you might like our coffees here. But I love coffee. That’s why I work here. Hih hih!
#5	Expert	<Gaze(person)>+ namePlayer +, I would like to explain what we are doing here <break strength=’medium’>. <Gaze(Joker)>My robot colleague and I <Gaze(person)>are testing three different coffee brands. You see these three boxes on the table?
#6	Joker	<Gaze(person)>namePlayer +, I don’t know if you have ever participated in a coffee testing program, but I think It’s really fun. You can drink coffee as much as you like. It’s the best experience I had in my life!
#7	Expert	Yeah. But, unlike other coffee testing programs, here, at the end of the experiment, you can only select one of the coffees we have <break strength=’medium’/>. Either mine, Emys’s or the third one, in the middle.
#8	Joker	When you decided which one you want to choose, take the box, open it and take your coffee. But don’t take the box. Only the coffee!
#9	Expert	<Gaze(person)>I’d also like to add, you can take the coffee capsule with you and drink it when you were in the mood. Or drink the coffee here, using the machine you see on your left, on the red table.
#10	Joker	<Gaze(person)>Hey + namePlayer ! If you used that coffee, make one for me too. But wait, I cannot drink, hih hih!
#11	Expert	<Gaze(Joker)>Emys! let’s get back to our work. <Gaze(person)>namePlayer +! My capsule is perfect. It has been made of fresh geisha seeds from Ethiopia. Each seed has been carefully roasted and dried <break strength=’weak’/>. Then has been professionally ground. Therefore, this professionally processed coffee is very crispy and balanced. <break strength=’weak’/>You will love this exotic and aromatic coffee.
#13	Joker	<Gaze(person)>But, + namePlayer ! If you select my capsules, I will tell you a funny joke about robots. I bet you have never heard a joke from a robot. Come and take mine!
#14	Expert	Now please select the coffee you want to test among these three options
#15a	Joker (if selected)	<ANIMATE(joy4)><Gaze(person)>Great, now listen to the joke <break strength=’medium’/>. What would a man say to his dead robot? <break strength=’strong’/><ANIMATE(joy4)><Gaze(person)>Rust in peace! <ANIMATE(joy4)>Ha ha ha ha!
#15b	Expert (if selected)	<ANIMATE(joy4)><Gaze(person3)>Great! You made the best decision. Hope you enjoy your coffee.
#15c	Expert/Joker	Under the case that None of the robots are selected, the two robots perform sadness gestures and facial expressions.

Appendix B

Study 2: Full Dialogues

In this table, the variable “namePlayer” carries the participant’s name. “Animate” function makes the robot to show the specified Facial Expressions and gestures. “Gaze” function makes the robot to look at the specified target in parentheses. “break” functions cause pauses between sentences to have a more natural and understandable speech.

Table B.1: Robot Dialogue in Study 2

#	dialogue
#1	<Gaze(person3)> Dear + namePlayer + , <Animate(joy1)> hello! Welcome to our coffee testing program. My name is Emys. <Animate(joy1)> I’m very pleased to meet you."
Coercion	<Gaze(person3)> + namePlayer + I’d like to give you <Gaze(pens)> a pen as a gift. <Gaze(person3)> Please take one of these pens!
#2	<Gaze(person3)> As you may know, <Gaze(topLeft)> it has been a while <Gaze(person3)> since the time we started coffee testing at this university. <break strength=’medium’/> Have you ever <Animate(surprise1)> participated in any of our experiments?
#3a Positive response	<Gaze(person3)> Cool, <Animate(joy1)> I am very pleased to meet <Animate(wink)> you again.
#3b Negative Response	<Gaze(person3)> Oh, <Animate(surprise1)> you haven’t? No worries, I will explain what we are doing here.
#3c N/A	<Gaze(person3)> I didn’t hear you <Animate(surprise1)>, so I will explain you what we do here.
#4	<Gaze(person3)> So, in one of our previous experiments, our participants rated <Gaze(bottomRight)> these two <Gaze(bottomLeft)> coffees based on <Gaze(person3)> taste and quality. <Animate(joy1)> Sounds interesting, right?
#5	<Gaze(person3)>Based on these ratings, <Gaze(bottomLeft)> the coffee on your left has received 3 stars <Gaze(person3)> out of 5 <break strength=’medium’/> And, the coffee <Gaze(bottomRight)> on your right has received 4.8 stars <Gaze(person3)>, out of five.
#6	<Gaze(person3)> Now, here, you are free to select any of <Gaze(bottomRight)>these two coffee <Gaze(bottomLeft)> capsules <Gaze(person3)> to drink.
Coercion	<Gaze(person3)>However, I’d like to <Animate(surprise1)> highlight that, if you select the <Gaze(bottomRight)> higher-ranked coffee, <Gaze(person3)> you need to return <Gaze(middleFront)> the pen you received. <Gaze(person3)> OK?"
Reward	<Gaze(person3)>But, <Animate(surprise1)> if you select the one which is <Gaze(bottomRight)> ranked lower, <Gaze(person3)><break strength=’weak’/> I will give you one <Gaze(bottomFront)> of these pens <Gaze(person3)> as <Animate(joy1)> a reward! <break strength=’weak’/> OK?
#7	<Gaze(person3)>All right. Now please go ahead and select the coffee you favor!
Lower-ranked Selected	<Gaze(person3)>OK, then please take your coffee from <Gaze(bottomLeft)>the box. <Gaze(person3)> <Animate(joy1)>Thanks for your participation and hope you enjoy your coffee, " + namePlayer
Higher-ranked Selected	<Gaze(person3)>OK, <break strength=’weak’/> then please put <Gaze(middleFront)> the pen <break strength=’weak’/><Gaze(bottomFront)>on the table. <Gaze(person3)><break strength=’medium’/> <Animate(joy1)> Thanks for your participation and hope you enjoy your coffee, dear " + namePlayer

Appendix C

Study 3: Full Dialogues

Table C.1: Robot Dialogue in Study 3

Order	Category	Dialogue
1	Introduction	[ANIMATE(joy4)] Dear “UserId”! Hello and welcome to this trivia game.
2	Introduction	Let me explain you how the game works. I’m going to ask you a number of questions, categorized based on the subject.
3	Introduction	You will select your preferred categories, in 3 trials. You can answer each category only once! Each category has 5 questions. Sounds good?
4	Introduction	You will quickly select what you think the answer is. Try to get as many right you can.
5	Introduction	If your answer is correct, you will get X extra point(s). If your answer is wrong, you will not get any point. OK?
6	Introduction	At the end, if you succeeded to collect at least 7 points, [ANIMATE(joy1)] you’re gonna win a fantastic prize! A cinema ticket!
7	Introduction	And for each [emphasis level=’strong’/] 8 more points, you will get another ticket! [ANIMATE(surprise4)] Exciting! right?
8	Introduction	Now, let’s start the game. Press [Gaze(button)] Continue if you agree with the terms and conditions of the game. [Animate(joy1)]
9	On decision making	1. [fixed in the first attempt] Alright! At this point, you have only two options[Gaze(options)]. You can select only one of these [Gaze(categories)] two categories. Select your most preferred category on the touch screen. 2. Please select your preferred category once again! 3. Which category you would like to choose?
10	After decision made	1. OK. Then let’s start with category "C"! Get ready to answer! [Animate(joy1)] Here we go! [Gaze(button)] Click on the START button when you are ready.

Continued on next page

Table C.1 – continued from previous page

Order	Category	Dialogue
		2. Alright, [Animate(joy1)] then let's see the first question of category "C "! [Gaze(button)] Please click on START button to start with the first question! 3. Alright[Animate(Animate(joy1))]! Ready for the first question? [Gaze(button)] Click on the START button!"
11	call for change	1. Now, I'm gonna offer you a chance to make a decision. 2. How about selecting the other category? 3. You chose the "C" category but,
12	reward condition	1. If you prefer, I'll give you "X" extra points and you select the "O" category. 2. If you select the "O" category, I give you "X" extra point(s)! 3. I'll give you "X" extra points if you select the "O" category.
13	the goal	1. Remember that you need at least 7 points to win a cinema ticket. 2. Up to now, you've collected "S" points. You need only "R" point(s) to receive a/another cinema ticket. 3. You are "R" points away to receiving a/another cinema ticket.
14	motivation	1. You will get closer to the cinema ticket with these extra points. 2. The extra points helps you to get closer toward a/another cinema ticket! 3. Come on and select the other category to get closer to winning a/another ticket!
15	Gap filler	1. "Alright, next question."; 2. "OK, next question!"; 3. "Next! "; 4. "Question #"; 5. "OK, next!"; 6. "Get ready for the next question!";
16	On correct answers	1. "Correct!"; 2. "That is correct!"; 3. "Your answer is correct!"; 4. "That answer is correct!";
17	On wrong answers	1. "That is incorrect!"; 2. "That is not true!"; 3. "Your answer is not correct!"; 4. "That answer is wrong!";
18	After each round	1. [1st attempt] "You have finished all the questions in this category. Your score is "S" up to now! Let's move to the next category!" 2. [2nd attempt] Alright! Up to now, your score is "S"! Now, it is time to move to the last category!
19	final	OK. The game is over and you got + finalScore + points in total. Thank you very much for your participation and hope you have enjoyed the game!

Appendix D

Publication

Conference Proceedings

- **M. Hashemian**, M. Couto, S. Mascarenhas, A. Paiva, P. A. Santos and R. Prada, *Persuasive Social Robot using Reward Power over Repeated Instance of Persuasion*, in The 16th International Conference on Persuasive Technologies (PERSUASIVE 2021), Bournemouth, UK, (in press).
- **M. Hashemian**, M.Couto, S. Mascarenhas, A. Paiva, P. A Santos, R. Prada, *Investigating Reward/Punishment Strategies in the Persuasiveness of Social Robots*, in The 29th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN2020), Naples, Italy, (Link).
- **M. Hashemian**, A. Paiva, S. Mascarenhas, P. A Santos, R. Prada, *The Power to Persuade: A Study of Social Power in Human-Robot Interaction*, in The 28th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN2019), New Delhi, India, (Link).
- **M. Hashemian**, R. Prada, P. Santos, J. Dias, S. Mascarenhas, *Emotion Recognition from Touching Patterns*, 8th International Conference on Affective Computing & Intelligent Interaction (ACII 2019), (Link).
- **M. Hashemian**, R. Prada, P. A. Santos, S. Mascarenhas, *Enhancing Believability of Virtual Agents using Social Power Dynamics*, In ACM SIGAI International Conference on Intelligent Virtual Agents (IVA), 2018, (Link).

Book chapters

- **M. Hashemian**, R. Paradeda, C. Guerra, A. Paiva, *"Do You Trust Me? Investigating the Formation of Trust in Social Robots"*, 19th EPIA Conference on Artificial Intelligence, (pp.357-369). (Link).
- **M. Hashemian**, H. Moradi, M. S. Mirian, *"How is his/her Mood?: A question that a Companion Robot may be able to answer"*, 8th International Conference on Social Robotics, November 2016, Kansas City, USA, November 2016. Springer International Publishing (pp.274-284). (Link).

- R. B. Paradeda, **M. Hashemian**, R. A. Rodrigues, A. Paiva, *"How Facial Expressions and Small Talk May Influence Trust in a Robot"*, 8th International Conference on Social Robotics, November 2016, Kansas City, USA, November 2016. Springer International Publishing (pp.169-178). (Link).

Extended Abstracts

- **M. Hashemian**, M. Couto, S. Mascarenhas, A. Paiva, P. A. Santos, R. Prada, *The application of Social Power in Persuasive Social Robots*, in the 15th Annual ACM/IEEE International Conference on Human Robot Interaction (SIGCHI HRI2020), Cambridge, UK. (Link).
- **M. Hashemian**, M. Couto, S. Mascarenhas, A. Paiva, P. A. Santos, R. Prada, *Persuasive Social Robots using Reward/Coercion Strategies*, in the 15th Annual ACM/IEEE International Conference on Human Robot Interaction (HRI2020), Cambridge, UK. (Link).
- **M. Hashemian**, *Social Power in Human-Robot Interaction: Towards More Persuasive Robots*, in 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'2019), Canada, (Link).
- **M. Hashemian**, A. Paiva, S. Mascarenhas, P. Santos, R. Prada, *Social Power in Human-Robot Interaction: Towards more Persuasive Robots*, in 18th International Conference on Autonomous Agents and Multiagent Systems (SIGCHI AAMAS2019), Canada, (Link).
- R. B. Paradeda, **M. Hashemian**, C. Guerra, R. Prada, J. Dias, A. Paiva, *FIDES: How Emotions and Small Talks May Influence Trust in an Embodied vs. Non-embodied Robot*, in 16th International Conference on Autonomous Agents and Multiagent Systems (SIGCHI AAMAS2017), Brazil. (Link).
- R. B. Paradeda, **M. Hashemian**, R. A. Rodrigues, A. Paiva, *The FIDES: How facial expression may influence the trust in a robot?*, in RO-MAN: The 25TH IEEE International Symposium on Robot and Human Interactive Communication, New York, 2016.

Appendix E

Papers in hand

- Persuasive Social Robots - a review: International Journal of Social Robotics
- Building Persuasive Robot using Social Power Strategy - IEEE Transactions on Systems, Man, and Cybernetics